

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:06:26 ; Search time 59 Seconds
(without alignments)
4635.127 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575

Sequence: 1 MAASVDQLKQVEALKMEN.....SSQTQSPKRHSGLVLTVS 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14533	99.7	2843	1	RBHUAP	adenomatous polypo
2	13148	90.2	2845	2	I49505	adenomatous polypo
3	3508.5	24.1	2274	2	T30258	adenomatous polypo
4	1778.5	12.2	2416	2	T13825	adenomatous polypo
5	638.5	4.4	5327	2	T13564	microtubule-associ
6	518	3.6	2271	2	F90073	hypothetical prote
7	517	3.5	2232	2	T34434	hypothetical prote
8	516.5	3.5	1186	2	T23327	adenomatous polypo
9	514.5	3.5	1188	2	T23330	hypothetical prote
10	477	3.3	2774	2	A43359	microtubule-associ
11	452	3.1	3507	2	T34513	hypothetical prote
12	442	3.0	4776	2	E95206	cell wall surface
13	428.5	2.9	2722	2	T20532	hypothetical prote
14	426.5	2.9	2738	2	E88320	protein F07A11.6 [
15	419.5	2.9	1791	2	T02345	hypothetical prote
16	417	2.9	2888	2	I49477	alpha-A-crystallin
17	416	2.9	2938	2	T30249	cell proliferation
18	412.5	2.8	3488	2	T34418	hypothetical prote
19	410	2.8	2187	2	T30826	nascent polypeptid
20	409	2.8	2526	2	T20531	hypothetical prote
21	406	2.8	5170	2	T15348	hypothetical prote
22	405	2.8	3542	2	T42730	Bassoon protein -
23	396	2.7	6713	2	E89921	hypothetical prote
24	395	2.7	3924	2	S37431	ankyrin 2, neurona
25	392.5	2.7	3147	2	T18674	hypothetical prote
26	391	2.7	5105	2	T32650	hypothetical prote
27	387.5	2.7	3187	2	JCS837	364K Golgi complex
28	384	2.6	1459	2	T32271	hypothetical prote
29	382.5	2.6	2176	2	T13806	toucan gene protei

30	381	2.6	2364	2	A56577	microtubule-associ
31	380.5	2.6	4377	2	A55575	ankyrin 3, long sp
32	375.5	2.6	2248	2	A35938	profilagrin - hum
33	375.5	2.6	3562	2	A47171	chondroitin sulfat
34	375	2.6	3968	2	A44265	trithorax homolog
35	373	2.6	971	2	T19431	hypothetical prote
36	373	2.6	3256	2	A48666	cell proliferation
37	372.5	2.6	2253	2	T30336	nuclear/mitotic ap
38	372	2.6	2649	2	T51023	hypothetical prote
39	369.5	2.5	2464	1	ORM9P1	microtubule-associ
40	368.5	2.5	2453	2	S60254	nuclear receptor c
41	367.5	2.5	2346	2	T13829	tpx homolog - frul
42	365.5	2.5	1560	2	T02885	peroxisome prolife
43	365	2.5	3869	2	A48205	All-1 protein -GTE
44	364	2.5	2441	2	D71623	erythrocyte membra
45	364	2.5	2897	2	B48666	cell proliferation

ALIGNMENTS

RESULT 1

RBHUAP

adenomatous polyposis coli protein - human

N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1991 #sequence, revision 31-Dec-1991 #text_change 21-Jul-2000

C;Accession: A37261; B39658; A44928; A49319; I54271

R;Kinzie, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith, chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A;Title: Identification of FAP locus genes from chromosome 5q21.

A;Reference number: A37261; MUID:91335210; PMID:1651562

A;Accession: A37261

A;Molecule type: mRNA

A;Residues: 1-2843 <KIN>

A;Cross-references: GB:W74088; NID:G182396; PIDN:AAA03586.1; PID:G182397

R;Joslyn, G.; Carlson, M.; Thilaveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grode

arrington, J.; McPherson, J.; Wasmuth, J.; Le Faslier, D.; Abderrahim, H.; Cohen, D.; Leif

Cell 66, 601-613, 1991

A;Title: Identification of deletion mutations and three new genes at the familial polypos

A;Reference number: A39658; MUID:91330307; PMID:1678319

A;Accession: B39658

A;Molecule type: DNA

A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P',

A;Cross-references: GB:W73548; NID:G190183; PIDN:AAA60354.1; PID:G190184

R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelste

Cancer Res. 52, 643-645, 1992

A;Title: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a cc

A;Reference number: A44928; MUID:92119623; PMID:1310068

A;Accession: A44928

A;Molecule type: DNA

A;Residues: 1506-1525 <MK>

A;Cross-references: GB:S78214; NID:G243541; PIDN:AA21145.1; PID:G243542

A;Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBI:P:78218)

R;Spurio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelbert

Cell 75, 951-957, 1993

A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.

A;Reference number: A49319; MUID:94073973; PMID:8252630

A;Accession: A49319

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 'G', 143-171, 'P', 173-179 <SPI>

A;Cross-references: GB:S67787; NID:G461061; PIDN:AAD13997.1; PID:G4261697

R;Lambertz, S.; Ballhausen, W.G.

Hum. Genet. 90, 650-652, 1993

A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly

A;Reference number: I54271; MUID:93186137; PMID:8383094

A;Accession: I54271

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-4 <LAM>

A;Cross-references: GB:S56365; NID:G256243; PIDN:AAD14918.1; PID:G4262770

C:Genetics:		Query Match		99.7%; Score 14533; DB 1; Length 2843;	
A:Gene: GDB:APC		A:Cross-references: GDB:119682; OMIM:175100		Best Local Similarity 99.8%; Pred. No. 0;	
A:Map position: 5q21-5q22		A:Note: mutations of this gene can result in familial adenomatous polyposis or sporadic		Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;	
C:Superfamily: adenomatous polyposis coli protein		C:Keywords: cancer; familial adenomatous polyposis; tumor suppressor			
F:1-730/Domain: leucine-rich <NTD>		F:1-730/Domain: leucine-rich <NTD>			
F:7-72/Region: coil #status predicted		F:7-72/Region: coil #status predicted			
F:185-227/Region: coil #status predicted		F:185-227/Region: coil #status predicted			
F:731-2832/Domain: serine-rich <CTD>		F:731-2832/Domain: serine-rich <CTD>			
F:1131-1156/Region: acidic		F:1131-1156/Region: acidic			
F:1558-1577/Region: acidic		F:1558-1577/Region: acidic			
F:1866-1893/Region: highly charged		F:1866-1893/Region: highly charged			
QY	1	MAAASYDQLLKQVEALKMENSLRQELDNLSNHLTKLETASNKKEVLKOLQGSIDEAM 60		Db	781
Db	1	MAAASYDQLLKQVEALKMENSLRQELDNLSNHLTKLETASNKKEVLKOLQGSIDEAM 60		QY	841
QY	61	ASSGQIDLLERLKLNDSSNFPQVKLRKMSLRSGSREGSVSSRGSCSPVPMGSFPR 120		Db	841
Db	61	ASSGQIDLLERLKLNDSSNFPQVKLRKMSLRSGSREGSVSSRGSCSPVPMGSFPR 120		QY	901
QY	121	RGFVNGSRESTGYLELEKERSULLADLKEEKEKWYLAQLQNLTKRIDSPLTENFSL 180		Db	901
Db	121	RGFVNGSRESTGYLELEKERSULLADLKEEKEKWYLAQLQNLTKRIDSPLTENFSL 180		QY	961
QY	181	QTDITRQLEVEARQIVAMEEQIGTQCDMEKQARRIARIQIIEKDIIRIQLLOSOAT 240		Db	961
Db	181	QTDITRQLEVEARQIVAMEEQIGTQCDMEKQARRIARIQIIEKDIIRIQLLOSOAT 240		QY	1021
QY	241	EAERSSQNKHETGSHDAERQEGGVGEINMATSNGNGQSTTRMDHETASVLSSSSTHSA 300		Db	1021
Db	241	EAERSSQNKHETGSHDAERQEGGVGEINMATSNGNGQSTTRMDHETASVLSSSSTHSA 300		QY	1081
QY	301	PRRLTSLGTVKVMYGLLSMLGTHDKDMSRILLAMSSQDSCISMRQSGCPLLIQLL 360		Db	1081
Db	301	PRRLTSLGTVKVMYGLLSMLGTHDKDMSRILLAMSSQDSCISMRQSGCPLLIQLL 360		QY	1141
QY	361	HGNDKDSVLLGNSRGSKEARASAAALNIIHSQPDQKRRRIIRVHLLEQLRAYCEIC 420		Db	1141
Db	361	HGNDKDSVLLGNSRGSKEARASAAALNIIHSQPDQKRRRIIRVHLLEQLRAYCEIC 420		QY	1201
QY	421	WEMQEAHEPGMDQKNMPAPVEHQICPVCVLMKLSFDEHRHANNELGGLQAIALLQ 480		Db	1201
Db	421	WEMQEAHEPGMDQKNMPAPVEHQICPVCVLMKLSFDEHRHANNELGGLQAIALLQ 480		QY	1261
QY	481	VDCSMYGLTNDHYITLRRVAGMALTNLTGPDVANKATLCSMGKCMRALVAOLKSESED 540		Db	1261
Db	481	VDCSMYGLTNDHYITLRRVAGMALTNLTGPDVANKATLCSMGKCMRALVAOLKSESED 540		QY	1321
QY	541	QQVIASVLRNLSRADVNSKKTREVGSVKALMECALVKKSTLKSIVLSALNLSAHT 600		Db	1321
Db	541	QQVIASVLRNLSRADVNSKKTREVGSVKALMECALVKKSTLKSIVLSALNLSAHT 600		QY	1381
QY	601	ENKADICAVDGAFLVGLTITYSQNTLAIIESGGGILRNVSLLIATNEDHQILRENN 660		Db	1381
Db	601	ENKADICAVDGAFLVGLTITYSQNTLAIIESGGGILRNVSLLIATNEDHQILRENN 660		QY	1441
QY	661	CLQTLQLKSHSLTIVSNACGLTNLNSARNPKDQDALWDMGAVSMLKNLIHSKHWIAM 720		Db	1441
Db	661	CLQTLQLKSHSLTIVSNACGLTNLNSARNPKDQDALWDMGAVSMLKNLIHSKHWIAM 720		QY	1501
QY	721	GSAALRNLMANRPKYKDNIMSPGSSLSPLHVRKQKALELDAQHSETFDNIDNLS 780		Db	1501
Db	721	GSAALRNLMANRPKYKDNIMSPGSSLSPLHVRKQKALELDAQHSETFDNIDNLS 780		QY	1561
QY	781	PKASHRSKQRHKQSLYGDFVFTNRRHDDNRSDNFTGNMTVLSPLYLNTTVLPSSSSSRGS 840		Db	1561

1921 GQPKPILQKQTFPOSSKDIIDRGAATDEKLNFAIENTPVCFSHNSLSLSIDIDENN 1980
 Db GQPKPILQKQTFPOSSKDIIDRGAATDEKLNFAIENTPVCFSHNSLSLSIDIDENN 1980
 1981 NKNEPIKETPPSQGSPKQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDLQ 2040
 Db NKNEPIKETPPSQGSPKQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDLQ 2040
 2041 ECISAMPKPKKPSRLKGNKSHPRNMGILGEDLTLDLKDIOQPSHGLSPDSENF 2100
 Db ECISAMPKPKKPSRLKGNKSHPRNMGILGEDLTLDLKDIOQPSHGLSPDSENF 2100
 2101 WKAIQEGANSIVSSHQAAAACLSQASSDSDLSLSKSGISLSPHLPDDEKPEFT 2160
 Db WKAIQEGANSIVSSHQAAAACLSQASSDSDLSLSKSGISLSPHLPDDEKPEFT 2160
 2161 SNKGRIILKPGKSTLETKKIESBKIGKGGKVKYKSLITGKVRNSISISQMKQPLQAN 2220
 Db SNKGRIILKPGKSTLETKKIESBKIGKGGKVKYKSLITGKVRNSISISQMKQPLQAN 2220
 2221 MFSISGRGTMIHIPGVNRNSSSTSPVSKGPPKTPASKSPSEGTATTPRGAKPVS 2280
 Db MFSISGRGTMIHIPGVNRNSSSTSPVSKGPPKTPASKSPSEGTATTPRGAKPVS 2280
 2281 ELSVARQTSQIGSSKAPSRGSDSTPSRPAQPLSRPIQSPGRNSISISQMKQPLQAN 2340
 Db ELSVARQTSQIGSSKAPSRGSDSTPSRPAQPLSRPIQSPGRNSISISQMKQPLQAN 2340
 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGRQMSQNLTKQTLGSKNASSIPRSESASG 2400
 Db KLSQLPRTSSPTASTKSGSGKMSYTPGRQMSQNLTKQTLGSKNASSIPRSESASG 2400
 2401 LNQMNGNGANKKVELSRMSSTKSGSDRERFVLVRQSTFIKEAPSPILRRKLESA 2460
 Db LNQMNGNGANKKVELSRMSSTKSGSDRERFVLVRQSTFIKEAPSPILRRKLESA 2460
 2461 SFESLSPSRPASPTRSQATQTPVLSPLPDMSLTHSSVQAGWRKLPNLSPTTIEYNDG 2520
 Db SFESLSPSRPASPTRSQATQTPVLSPLPDMSLTHSSVQAGWRKLPNLSPTTIEYNDG 2520
 2521 RPAKRHDIAARSHSESPRLPINRGSTGWKREHKGSSSLPRVSTWRTTSGSSSILSASSES 2580
 Db RPAKRHDIAARSHSESPRLPINRGSTGWKREHKGSSSLPRVSTWRTTSGSSSILSASSES 2580
 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 Db SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 2641 KTLIYQMAVASKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKSDKN 2700
 Db KTLIYQMAVASKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKSDKN 2700
 2701 QAKQNVGNGSVPMRTVGLNRLTSPFOVADPOKGTETIKPGQNNPVVSETNESPIVER 2760
 Db QAKQNVGNGSVPMRTVGLNRLTSPFOVADPOKGTETIKPGQNNPVVSETNESPIVER 2760
 2761 PFSSSSSKHSSPSTVAARVTPFNPNPSRKSSADSTGARPSPQIPTPVNNNTKKRDSKT 2820
 Db PFSSSSSKHSSPSTVAARVTPFNPNPSRKSSADSTGARPSPQIPTPVNNNTKKRDSKT 2820
 2821 DSTESSGTQPKRHSGLVTSV 2843
 Db DSTESSGTQPKRHSGLVTSV 2843

RESULT 2

149505
 adenomatous polyposis coli protein - mouse
 N;Alternate names: APC
 C;Species: Mus musculus (house mouse)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C;Accession: I49505

R;Su, L. Science 256, 668-670, 1992
 A;Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
 A;Reference number: I49505; MUID:92263101; PMID:1350108
 A;Accession: I49505
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-2845 <RES>
 A;Cross-references: GB:M88127; NID:gl91991; PIDN:AB59632.1; PID:gl91992
 C;Superfamily: adenomatous polyposis coli protein

Query Match 90.2%; Score 13148; DB 2; Length 2845;
 Best Local Similarity 90.0%; Pred. No. 0;
 Matches 2566; Conservative 113; Mismatches 159; Indels 12; Gaps 10;

QY 1 MAASVDQLLKQVEALKMENSLRQLEDNNSHLTKLETEASNKKEVLKOLQGISIDEAM 60
 Db 1 MAASVDQLLKQVEALKMENSLRQLEDNNSHLTKLETEASNKKEVLKOLQGISIDEAM 60
 QY 61 ASSQIDLLERLXELNLDSSNPFVKLSRQMSLSRSGSVSSRSGECPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEFNLD-SNPFVKLSRQMSLSRSGSVSSRSGECPVPMGSPFR 118
 QY 121 RGFVNGSRESTGYLELEKEKERSILLADLDKEEKEKQWYQAOLNQLKRIIDSLPTENFSL 180
 Db 119 RTFVNGSRESTGYLELEKEKERSILLADLDKEEKEKQWYQAOLNQLKRIIDSLPTENFSL 178
 QY 181 QTDLTRELEVEARQTEVAMEEQLGTCODMEKRAORRIARIQOIEKDILRIOLLOSOAT 240
 Db 179 QTDLTRELEVEARQTEVAMEEQLGTCODMEKRAORRIARIQOIEKDILRIOLLOSOAA 238
 QY 241 EAERSSQKHETGSHDAERQNEGGVGEINMATSGNGQSTTRMDHETAVLSSSSTHSA 300
 Db 239 EAERSSQSRHDAASHEAGRQHEGHGVAESNTAASSSQSPATRVDTETASVSSSGTHSA 298
 QY 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 360
 Db 299 PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMRQSGCLPLLIQLL 358
 QY 361 HGNDKDSVLLGNRSGSKEARASAAALHNIHSQPDCKRGRREIRVLLLEQIRAYCETC 420
 Db 359 HGNDKDSVLLGNRSGSKEARASAAALHNIHSQPDCKRGRREIRVLLLEQIRAYCETC 418
 QY 421 WEWOEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRHANNELGLOIAIELQ 480
 Db 419 WEWOEAHEQGDQDNMPAPVEHQICPAVCVLMKLSFDEHRHANNELGLOIAIELQ 478
 QY 481 VDCWYGLTNDHYGITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESED 540
 Db 479 VDCWYGLTNDHYGITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSESED 538
 QY 541 QQVIASVLRNLWRADVNSKKTLEEVGSVKALMECALEVKKESTLKSVALMNLSAHCT 600
 Db 539 QQVIASVLRNLWRADVNSKKTLEEVGSVKALMECALEVKKESTLKSVALMNLSAHCT 598
 QY 601 ENKADI CAVDGALAFVGLTLYRSQTNLTALIESGGGILRVNSSLIATNEDHROILRENN 660
 Db 599 ENKADICAVDGAFLVGLTLYRSQTNLTALIESGGGILRVNSSLIATNEDHROILRENN 658
 QY 661 CIQTLQLHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGVSMKLNLIHSHKMIAM 720
 Db 659 CIQTLQLHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGVSMKLNLIHSHKMIAM 718
 QY 721 GSAAALRNLMANRPKAYKDANIMSPGSSLPSLHVKKOKALEAELDAQHLSSTFENIDNLS 780
 Db 719 GSAAALRNLMANRPKAYKDANIMSPGSSLPSLHVKKOKALEAELDAQHLSSTFENIDNLS 778
 QY 781 PKASHRSKORHKQSLYGDYVDFDTNRHDDNRSDNFNTGNMTVLSPLYNTVLPSSSSSRGS 840
 Db 779 PKASHRSKORHKQSLYGDYVDFDTNRHDDNRSDNFNTGNMTVLSPLYNTVLPSSSSSRGS 838
 QY 841 LDSSSEKDRSLERERGIGLGNYPHATENPGTSSKRGLOISTTAAQIAKAKMEVSAIHTS 900

Db 839 LDSSRSKDRSLERBERGICLSAYHPTTENVAGTSSKRGIGIQTITTAQAIAKVMEEVSAIHTS 898
Qy 901 QEDRSSGSTEHLCHVTDERNALRRSSAAHTSNVNTFTKSENSNRTCSMPYAKLEYKRSS 960
Db 899 QDRSSASTTEPHCVADDRSAAARSASHTSNVNTFTKSENSNRTCSMPYAKVEYKRSS 958
Qy 961 NDSLNSVSNQGYKRGQWKPSIESYSDDESDFCSYGYCPADLAHKIHSANHMDNDGE 1020
Db 959 NDSLNSVTSOGYKRGQWKPSVSYSDDESDFCSYGYCPADLAHKIHSANHMDNDGE 1018
Qy 1021 LDTPINYLKYSDEOLNSGRSPQNERWARPKHIIEDIEIKSQSORQSRNQSTTYPVYTE 1080
Db 1019 LDTPINYLKYSDEOLNSGRSPQNERWARPKHIIEDIEIKONEORQARSQNTSYPVYSE 1078
Qy 1081 STDDKHLKFQFHGQOECVSPYRSGANGSETNVRGSHNGINQNVOSLCOEDDYEDDKP 1140
Db 1079 NTDDKHLKFQFHGQOECVSPYRSGGTSGSTNRGSSHAINQNVOSLCOEDDYEDDKP 1138
Qy 1141 TNYSERVSEEPQH-EEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKSPFSFK 1199
Db 1139 TNYSERVSEESQHEEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKSPFSFK 1198
Qy 1200 SSSQSSKTEHWSSESNTSTPSSNAKQNLHPSSAQSGQDQKAACTCKVSSINQETI 1259
Db 1199 NSSAQSTKPEHLSPESENTAVPPSNAKRQNLRFSSAQ-RNGQTKGTTCKVPSINQETI 1257
Qy 1260 QTYCVEDTPICFRSCSSLSLSSAEDEIGCQNTTQEADSANTLQIABIKGKITRGAEDP 1319
Db 1258 QTYCVEDTPICFRSCSSLSLSSADDEIGCQNTTQEADSANTLQIABIKGKITRGAEDP 1317
Qy 1320 VSEVPAYSQHPTKSRKLOGSSLSSESARH-KAVEFPSSGAKSPKSGAQTPKPPBHYVQ 1378
Db 1318 ATEVPAYSQHPTKSRKLOGSSLSSESARH-KAVEFPSSGAKSPKSGAQTPKPPBHYVQ 1377
Qy 1379 ETPLMFRCTSVSLDFEBSRIASSVQSPBCSMVSGIISPDLDPSPQCTMPPRSKRT 1438
Db 1378 ETPLMFRCTSVSLDFEBSRIASSVQSPBCSMVSGIISPDLDPSPQCTMPPRSKRT 1437
Qy 1439 PPPPQTAQTRKVPKNAKPAEKRESGPKQAAVNAAVQVQLPADTLLHFEATESTPD 1498
Db 1438 PPPPQTAQTRKVPKNAKPAEKRESGPKQAAVNAAVQVQLPADTLLHFEATESTPD 1497
Qy 1499 GFSCSSLSLSALDELPEFIQKDVLRINPPVQENDNGNETESEPESNENQEKAKTID 1558
Db 1498 GFSCSSLSLSALDELPEFIQKDVLRINPPVQENDNGNETESEPESNENQEKAKTID 1556
Qy 1559 SEKDLDDSDDDDIIELEECIIISAMPTKSRKGGKPAQATSKLPPVPARKPSQLPVYKLL 1618
Db 1557 SEKDLDDSDDDDIIELEECIIISAMPTKSRKGGKPAQATSKLPPVPARKPSQLPVYKLL 1616
Qy 1619 PSQNLQPKHVSFTPGDDMPRYVCVEGTPIINFSTATSLDITIEPPEHLAAGSVRG 1678
Db 1617 PAQNLQPKHVSFTPGDDMPRYVCVEGTPIINFSTATSLDITIEPPEHLAAGSVRG 1676
Qy 1679 AQGEFEKRTDPTIEGRSTDEAOGKTSVTIPELDDNKAEGDIIAECINSAMPKGS 1738
Db 1677 IQGEFEKRTDPTIEGRSTDEAOGKTSVTIPELDDNKAEGDIIAECINSAMPKGS 1736
Qy 1739 KPRVKKIMQVQAASASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRNADSKN 1798
Db 1737 KPRVKKIMQVQAASASSAPKNQLDGKKKPTSPVKPIPONTYRTRVRNADSKN 1796
Qy 1799 LNAERFSDNKKQKQNLKNSXDFNDKLPNNEDVRGSPFDPSPHYTPIEGTPCFER 1858
Db 1797 VNTTEFPDNDKSKPSLQNAKAFNEKLPNNEDVRGTFALDSPHYTPIEGTPCFER 1856
Qy 1859 NDSLSDLDFODDVDVLSREKAEKRLAKENKESPAKVTSHTELTSNOQSANKTCAIAKQPI 1918
Db 1857 NDSLSDLDFODDVDVLSREKAEKRLAKENKESPAKVTSHTELTSNOQSANKTCAIAKQPI 1916
Qy 1919 NRQOKPILQKSTFPQSSKIDPRGAATDEKLQNEFAIENTPVCFSHNSLSLSDIOE 1978
Db 1917 NRAQSPVLQKQTFPQSSKIDPRGAATDEKLQNEFAIENTPVCFSHNSLSLSDIOE 1976

Qy 1979 -NNNKENPIKETEPDPSQGEPSKPAQGYAPKSFHVEDTPVCFSRNSSLSSLSIDSDD 2037
Db 1977 NNNKSEPIKEAPEANQGEPSKPAQGYAPKSFHVEDTPVCFSRNSSLSSLSIDSDD 2036
Qy 2038 LLOECISSAMPKKKPSRLKGNKHSRNMCGILGEDLTLDLKQIQRDSEHGLSPSE 2097
Db 2037 LLOECISSAMPKKKPSRLKSEKSPKVGILAEEDTLDLKQIQRDSEHGLSPSE 2096
Qy 2098 NFDWKAIOGANSIVSSLHQ-AAAAACLSRQASSDSLSILSKSGISLGSPPHLPDQEE 2156
Db 2097 NFDWKAIOGANSIVSSLHQ-AAAAACLSRQASSDSLSILSKSGISLGSPPHLPDQEE 2156
Qy 2157 KPFTSNKGRPILKPGKSTLETKTIESKSGIKGKGVYKSLITOKVRSNBEISQMKQP 2216
Db 2157 KPFTSNKGRPILKPGKSTLEAKTIESKSGIKGKGVYKSLITOKVRSNBEISQMKQP 2216
Qy 2217 LOANMPSISRGRTHIHPGVNRNSSSTSPVSKKGPPLKTPASKPSEGOATTSPRGAKP 2276
Db 2217 LETNPFISRGRTHIHPGLNSSSTSPVSKKGPPLKTPASKPSEGOATTSPRGAKP 2276
Qy 2277 SVKSELSPVAROTSOIGGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGRNGI 2336
Db 2277 AGKSELSPITROTSTIGSNKSGSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISPGRNGI 2336
Qy 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSTSPGRMSQOQLTKOTGLSKNASSIPRSES 2396
Db 2337 SPNKLSQLPRTSSPSTASTKSSGSKMSTSPGRMSQOQLTKOTGLSKNASSIPRSES 2396
Qy 2397 ASKGLNOMNMGANKKVELSRMSTKSSGESDRSERPVLVROSTFIKEAPSTLRKL 2456
Db 2397 ASKGLNOMNMGANKKVELSRMSTKSSGESDRSERPVLVROSTFIKEAPSTLRKL 2456
Qy 2457 BESASFESLSPSSRPASPTRSOACTPVLSPSLPMSLSTHSSVOAGGWRKLPNLSPTIE 2516
Db 2457 BESASFESLSPSSRPASPTRSOACTPVLSPSLPMSLSTHSSVOAGGWRKLPNLSPTIE 2516
Qy 2517 YNDGRPAKRHDIAKSHSPSRRLPINRSGTWKREHSHSSSLPRVSTWRTTSSSSSILSA 2576
Db 2517 YNDGRPAKRHDIAKSHSPSRRLPINRSGTWKREHSHSSSLPRVSTWRTTSSSSSILSA 2576
Qy 2577 SSESSEKAKSEDEKXVANSISGTQSKENQVSAKTWIKENEFSPNTNSTQTVSSGATN 2636
Db 2577 SSESSEKAKSEDEKXVANSISGTQSKENQVSAKTWIKENEFSPNTNSTQTVSSGATN 2636
Qy 2637 GAESKTLIYQMAPAVSKTDEYVWRIEDCPIINPRSGRPTGNTPPVIDSVSEKANPIKD 2696
Db 2637 GAESKTLIYQMAPAVSKTDEYVWRIEDCPIINPRSGRPTGNTPPVIDSVSEKANPIKD 2696
Qy 2697 SKDNQ---AKONVNGSVPMRTVGLNRLTSFIOVDAPDQKGTIEIKGQNNPVPVSETNE 2753
Db 2697 SKDNQ---AKONVNGSVPMRTVGLNRLTSFIOVDAPDQKGTIEIKGQNNPVPVSETNE 2753
Qy 2754 SPIVERTPFSSSSSSSKHSPSGTVAARTPNYNPSPKSSADSTARPSCQIPTPVNNNT 2813
Db 2754 SPIVERTPFSSSSSSSKHSPSGTVAARTPNYNPSPKSSADSTARPSCQIPTPVNNNT 2813
Qy 2814 KKRDSKTDSTESSGTQPKRHSGLVTSV 2843
Db 2814 KKRDSKTDSTESSGTQPKRHSGLVTSV 2843

RESULT 3

T30258

adenomatous polyposis coli protein 2 - mouse

N/Alternate names: APC2 protein

C/Species: Mus musculus (house mouse)

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C/Accession: T30258

R/van Es, J.H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Miles, A.; Kuipers, J.

Curr. Biol. 9, 105-108, 1999

A/titile: Identification of APC2, a homologue of the adenomatous polyposis coli tumour sui

A/Reference number: 220796; MUID:99147086; PMID:10021369

A:Accession: T30258
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2274 <VAN>
A:Cross-references: EMBL:AJ130783; NID:G4210431; PIDN:CAA10207.1; PID:G4210432
C:Genetics:
A:Gene: APC2
A:Introns: 47/3; 78/1; 138/2; 174/3; 212/3; 238/3; 271/3; 396/1; 428/1; 474/3; 500/3; 53

Query Match 24.1%; Score 3508.5; DB 2; Length 2274;
Best Local Similarity 34.4%; Pred. No. 1.7e-136;
Matches 1005; Conservative 356; Mismatches 789; Indels 773; Gaps 98;

QY 4 ASYDOLLKQVAKMNSLNQCELDNSNHLTKLETSANMKEVLKOLQGSIEDEA--MA 61
DB 6 ASYEQLVQVQKALKANTHLQELRDNSHLSKLETETSGNKEVLKHLQGLKEARVIV 65
QY 62 SSGQIDLLERLKLNLDSNPFVGLRSLRSYSGSREGSVSRSGSCSPVPMGSPFRR 121
DB 66 SSGQTEVLEQLKALQTDISSLYNLKFHAP---ALGPEP---AARTPEGSPV-HGSGPSK 117
QY 122 -GFVNGSRESTGYLEBEKERSLLADLDKEKEDWYVYQALQNLTKEDSLPTENFSL 180
DB 118 DSFGLSRATIRLLBELDQRCFLLSIEKEKEKLYYSQQLSKLDELPHVDTFSM 177
QY 181 QDILTRROLEVBAROIRVAMEBQLGTCODMEKRAQRRIARIQOIEKDIL----RIQ--- 233
DB 178 QNDLIRQOLEFAQHIRSLMEBFQTSDEMVOQAIRASRLQEIIDKELLEAQDRVQOQTEP 237
QY 234 --LLOSQATEARRSQNHETGSHDAERQNEQGVGEINMATSNGQSGSTTRMDHETASV 291
DB 238 QALLAVKPVAVEEQEAEVTHPEDPGTPQ-----PGN----- 269
QY 292 LSSSTHSAPRLTSHLGHKVMYVSLLSMLGTHDKDMSRTLLAMSSQSDSCISMROSG 351
DB 270 -----SKVEVFWLLSNLARTROEDTARTLLAMSSPSCVAMRSG 311
QY 352 CLPLLIQLLHNDKDSV---LLGNRSGSKAPARASAAHLNIIHQOPDDKRRIRVLH 408
DB 312 CLPLLIQLLHGTAGSVGRAGIPGAPGAKAMRANAALHNIVFQDQGLARKEMRVLH 371
QY 409 LLEQIRAYCETCEWCEAEHPGMDQKNPMPAPVHOICPAVCLMKLSFDEEHRANNE 468
DB 372 VLEQIRAYCETCWDMLQASDGTG-----TPVPIEQICQATCAVNLKLSFDEEHRANNE 426
QY 469 LGLQIAIAELLQVDCMYGLTNDHYSITLRRYAGMALTNLTGDEVANKATLCMSKGMRA 528
DB 427 LGLQIAIAELLQVDEYHMKWTRDPLNLALRRYAGMTLTNLTFGDVANKATLCARRGMEA 486
QY 529 LVAQLKSEDIQOVIASVLRNLSWRADVNSKKTLEVGSVKALMECALEVKESTLKV 588
DB 487 IVAQLKSEDEELHQVVSILRNLWRADINSKVLREVGSMALMECVLRASKESTLKV 546
QY 589 LSAWNLASHCTENKADICAVDGAFLVGTLYRSQTNLTALIESGGGILRNVSLLIAT 648
DB 547 LSAWNLASHCTENKAAICQVDGALGFLVSTLYRCQGNLSLVIESGGGILRNVSLLIAT 606
QY 649 NEDHQILRENNCLQTLLOHLKSHSITIVSNAGCTLWNLISARNPKDQEAALMDGAVMLK 708
DB 607 REDYRQVLRDNLQTLLOHLKSHSITIVSNAGCTLWNLISARSPPRQOELLMDGAVMLR 666
QY 709 NLTHSKHMTAMGSAALNLMANRPKYKDANI--NSPGSSPLSHVRKOKALELDAQ 767
DB 667 NLVHSKHMTAMGSAALNLMANRPKYQAAVAVSPGTCVPSLVYRQRALAEALDTR 726
QY 768 HLSETFDNIDNLS--PKASHRSKQ-----RHQSLYGDYVFDNTRHDDNSDNFNTGMT- 820
DB 727 HLVAHLGHLKQSLPEAETTSKKPLPLRLHLDGLVDQVYASDSCGCFDDDDAPSAAAAATTA 786
QY 821 -----VLSPLYNTVLPSSSSRSGSLDSSRSKDRSLEREGIGLGNVHPATENPGTSS 874
DB 787 EPASPATVMSFLGPPFLQOALAR-----TPPARQGLGEAREAG-----GEAAVAA 833

875 KRGLOISTTAAQIAKVMEEVSAIHTSQEDRSSGSTITELHCVTDERNALRSSAAHTSNT 934
DB 834 KAKAKALAVARIDRLVEDISALHTSSDSDSFLSS-----GDPGQEAPEGRAQSCSPC 887
QY 935 YNFT--KSENSNPTCMPYAKLEYKRSNDNSNSVSDNGYKRGQMKPISIESYEDDES 992
DB 888 RGTEGGRREAGSRAHPLRLKKAHTSLNSDLSNGSTSDGYCTREHMTF----- 936
QY 993 KFCYSYGOYPADLAHLKHSANHMDNDGELDTPIYNSLKYSDQLNSGRGSPSONERWARP 1052
DB 937 --CPLAAL-----AHRDD-----PYRGQ--TRP 956
QY 1053 KHIIIEDEIKOSEQRORNOSTTVVYVTESTDDKHLKFPHQGOECVSPYRSRGANGSET 1112
DB 957 RLRLDPLGRAELPARDTAATDARVT-----IKLSPTYQHVPLLD-----GAAGA-- 1002
QY 1113 NRVGSHNINQNVQSQCQEDDDVEDDKPTINYSERYSEEQHEEBEERTNYSIKYNEEKRH 1172
DB 1003 -----GVRLPVPGTGS----- 1013
QY 1173 VDQPIDYSLKYATDIPSSQKQSFSPKSSGQSKTEHMSSESSENTSTPSSNAKRONQLH 1232
DB 1014 -----FGARKQAM----- 1021
QY 1233 PSSAQSSGQPOKAAATCKVSSINOETIOIYCVEDTPICFSRCSLSLSSAEDEICGNQT 1292
DB 1022 -IPADLSLKYPEKLIVASPL-PIASKVLQKLVADQGMWSGRCSSLSLSSSTGHAV--PSQ 1077
QY 1293 TOEADSAANTLQIAEIKG---KIG-----TRSAEDPVSEVPVAVSQHPRTKSSRLQSSLS 1343
DB 1078 AENLSDSSLEGLEAGPGEALGRAWRASGTSFLVS--IPA---PQRGRSR---GLG 1128
QY 1344 SESARHKAVFPFGAKSPSGAGCTPKSPPEHYVQETPLMFPSRCTSVSSLDSPESRSIAS 1403
DB 1129 VEDA-----TPSSSENVCQETPLVLSRCSVSSLSGSPESRSIAS 1168
QY 1404 SVQSEPCSGMVGIIPSPDLPSFGQTMPPSRKTPPPPPQTAQTKR-----EVPKNKA 1457
DB 1169 SIPDPCSGLSGTVSPSELDPFGQTMPPSRKTPPPPPQTAQTKR-----EVPKNKA 1228
QY 1458 PTASKRESGPKAAVAAVQVLPDADTLHLFATESTDPGFSRSSLSLSDLEPFIQ 1517
DB 1229 DIADCRERQPPSELDAQSVR-----FTVEKPDENFSCASSLSALALHLYVQ 1276
QY 1518 KDVELRMPVPV-QENDNGNETESEQPKESNENKEAEKTIIDSEK---DLLDDSDDDDDIE 1573
DB 1277 QDVELRLRPPACPRAYVG-----GGHRRDEAASRLDGPAPAGRSARSATDEKELE 1327
QY 1574 ILEECIISAMPTKSRKPKPAQTAASKLPPPVARKPSQLPVYKLLPSQNLQPKHVSFT 1633
DB 1328 ALRECLGAAMPARLK-----VASALVP--GRSLPVVYMLVPAPAR----- 1368
QY 1634 PGDDMPVYCVGTPINFSTATSLDITISPPNELAAGEVGRGGAQSGEPEKEDTIPT 1693
DB 1369 -GDD--SGTDSAEPTGVNFSSAASLSDTELQGPSRDKPAGPGR-----OK-----PT- 1413
QY 1694 GRSTDEAGGGTSSVTPIELDDNKAEBGDI LAECINGAMPKSHKPFVKKIMDQVQA 1753
DB 1414 GRAPARO-----TRSRP-----KAA 1430
QY 1754 SASSAPNKLQDGGKKKPTSPVKPIQNTY-RTVRKXNADSKNNLNAERVFSDNDSK 1812
DB 1431 GAGKS-----TEHTRGCRNRAGLEPLSRPOSASRNRDS 1466
QY 1813 KQNLKNSKDFNDKXLPNNEDVRGSAFDPSPHHVTPTEGTPYCFSRNDSLSLSDFDODDV 1872
DB 1467 COT-----RTRGGALQSLCLTPTPEEAYCF-----YDSDEE 1499
QY 1873 DLS-----RBKAEIRKAKENKSEAKVTSHTSLTSNOQSANKTQAIKQPIN---RQGP 1923
DB 1500 PPATAPPFRASALPRALKREKPAKGRK-----ETPSRAAQATPLFVRAQP 1544
QY 1924 KPILQKQSTFFQSSKDIIDPRGAATDEKLNFAIENTPVCFSHNSLSLSDIDQENNKE 1983

770 -----ENATKTSY----- 778
1110 SETNRVGNHGNVNSQLCQEDDYEDDKPTNYSERYSEREEBERPTNYGIKNEE 1169
779 -----QETDL--DQPTDSLRYAE-----NQIESDLDISGPAGG 811
1170 KRHVDPIDYSLKVATDIP--SQKQSFPSKSSGQSKTSH-----MSSSENSTPSS 1223
812 KSTITPP-----AETVPESEGEIILLIILDDSVKCYQEDTTPYVISNAASVTDLRVAA 864
1224 NAKRQNLHPS--SAQSRSGOPO--KAATCKVSSINOETIQTVCVEDTPIPCSRCSLS 1278
865 KADRAEVKPEVREVTSEKAPKPLKPLKUSQCGSGSYTPEKINCEGTGPFYFRYDLS 924
1279 SLSSAEDRIG--CNQTOFADRSANTLQIAEIKIGKIGTSAEDPVPVAVSQHPRTKSSRL 1337
925 SL-----DESGRANQAIUGTD-----ADIKPLEKQEQE----- 954
1338 QGSSLSSESARHKAVERFPSSGAKSPKSGAQTPKPPEHYVOTPLMBSRCTSVSLSLDFE 1397
955 -----SOPAEQVLTKPPTQANS-----ALETPLMFSRRSSMDSLVHP 992
1398 SRSIA-----SSVQSEPCSGMVSGIISPSDLDPDPSGQTMPPSRSKTTPPPPTQATKREV 1452
993 DVDVANCDDKSSVSD--FSRLASGVISPEIPDSTQSMPOS----- 1033
1453 PKNKAPTAKRESGPKQAANVAVORVQVLPDADTLHFATETPDGSCSSLSALSLD 1512
1034 PRNSVAGSGQNDVPPVVPVIFASLOPLRSVFE--DDLSSFNVEHTPAQFSTATISLSLI- 1091
1513 BPFIQKVELRIMPVQENDNGNE-----TESEQPKESNENQKE--AEKTIID 1558
1092 -----VDDEKAPAVWTEDEDELALLANCINGMQRKPTAVKSTVNSVDVAETIR 1144
1559 S-----EKD-----LLDD--SDDDDI----- 1572
1145 SYCTEDTPALLSKVPSNTNLSVISMSSTDPKDATAGQAQMAHQLSDDVSSNASDCGAS 1204
1573 --EILECIIISAMPTKSRGKGPATASKLPPPVARK--PSQLPVYKLLPSONR--QPQKHV 1630
1205 GHLLQOQIRDM-----KKPLGEATSDPILMURRGNELEFY--LPS----- 1244
1631 SFTPGDDPRVYCVBGTPTINSTATSLDLATIES-----PPNELAAGEVGRGAOSGEFEK 1686
1245 ---ADEMNK--FLVEDSPCNFVSWSGLNLTGVSSLVGPVQLKETE--PSSADQNPENMR 1298
1687 RDTITEGRSTDEAOGKTSSTVITPELDNKAEBGDILAE----- 1727
1299 SLANKSKRRPPHWQDDSLSSLSIDSEDDTNLLSQAIAGCNRPKSNLGFSSNGKRSSL 1358
1728 -----INSAMPKGGKHPFRVKIMQVQOASASSAPKNQO-----LDG--KK 1769
1359 SSSQPIAIAATSSLSNLSAMTVRSQQOESYSSVSDSDSDNDNQSKSLFELCLKGMYKT 1418
1770 KKP-----TSPVKPIPONTTEYRT-----RVRKNADSKNLNAER 1803
1419 KEFGARAQOMQBPVIGSSSVQSNPSLQFSLPVLQPSGGQVKQRHHHHHHHRRER 1478
1804 VFSDNKOSK--KQNLKXN--SKDFNDKLPNNEDRVKGSFAFDSPHYTPIEGTPTCYFSRND 1860
1479 -----ERKDEKLQECINTGLSKIN--AVPKXV--LATSAAALEPCH--PMAAT-----T 1523
1861 SLSSLDFFDDDDVDLGRKAEALRKAKENKESAKVTSHTELTNSQOASANKTOA--IAQPI- 1918
1524 SASALSTAAPDVE---QKAH---ATSNPQOQSSSTHPSSHILNPIDAIATVTDTRSPAA 1577
1919 -NRGQPKPILOQSTPPQSKDIPRGAATDEKLQNFALN-----TPVCF----- 1963
1578 PNQGNAGNSQNGLET--ATGSKOLDSEDRSSDESQSFIMETWRLDLSALNETCISGASEK 1636
1964 -----SHNSLSLSLSDIDQNNKNEPIKETEPDPS 1995
1637 HKDPDLMLKSLVERLTWEFTVSAEQLRSSSHNHSS-----NSHKKNSSNNTWNESTCFND 1691

RESULT 5

Ti3564

microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)

N:Alternate names: hypothetical protein EG:49E4.1

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13564

R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.

submitted to the EMBL Data Library, April 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: 217689

1996 QGEPSPQASGVAPKSFHVEDTPVCFSRNSSLSLSID---SEDDLLQECISAMPKKKK 2052
1692 VSPFVSQTA-----PV-----LASLDEDATEARSLHELIELITPTNEQQ 1732
2053 PSLKGDNEK---HSPRNMGILGEDTLIDLKDIOQPSEHGLSPDSENFOWKATCEGA 2108
1733 PESLEGETDTLVNGHADSYSGSGGLNFOL---GGQVQAGVRLEPQRLLENFTSASIMT 1789
2109 NSTVSSLIHQAAAAACLSRQASDSD-----SILSL-----KSGI--SLGSPF--HLTP 2152
1790 NSTMIAREALAEALLQPAATDDDDTTFTSLNSLDLDIRPPSGMESLNSCYQDHSOP 1849
2153 DQEEKPFTSNKGPRIILKPEKSTLETYKLIBESKGIKGGKVKYKSLITKVRNSNISGQ 2212
1850 SSRQAMPS--KSPRFARKFPANLVARALHLAG-----SAESVNSSCNLLDN 1897
2213 MKQP-----LQANMPISIRGRTWIHIPGVNRNSSSTSPVSKKGPPL----- 2253
1898 IKPFLMDELDDLSMIVSDISIQSEVADGEQDCSWATTISVNYETAACDDQTMVLOSCFD 1957
2254 -----KTPASKSPSEGQATATTPRGAKEVSKSEL-----SPVAR 2287
1958 EDEDATNDYSSAESTPKHGSTPSPNRRSLTPKDKERLTAKDRPKTYTIATSCMEAPEAN 2017
2288 QTSOI-----CGSSKAPSRSGRSDSTPSRPAQQLSRPI----- 2321
2018 ETLOEIVEAANVPATPSPRANGRRGSAERYKTQIECPALIQPODDCPSEQLSSIR 2077
2322 -----QSP--GRNSISPRNG-----ISPPNKUS 2343
2078 AMMQQFTFITDINIGHSQETCESTDHPDAGESPECQDQNETSCDQEPDHLPPPSIV 2137
2344 QL-----PTSPSPSTASTKSSGSKMSYTSRQMSQQNLTKOTGLSKNASSIPRSESA 2397
2138 DLRTSVVKPTTLEPATAVKLVGRKKPVPVSPVSMQSRN----- 2177
2398 SKGLNMQNNGCANKVKVELSRMSTKS-----SGESDSESRPV-----LVRQSTP. 2443
2178 -----SNNAAASKKLTLPPTAKESLVPGSGVRLPAKKXTPTPEPAPALERQGT 2230
2444 IKRAPSPTLRKLEESASFESLSPSRPAPTQSOQATPVLSPLDMSLSTHSSVQAGG 2503
2231 VKD-----EPTNSNVQVPVVE--TKPAQTSPTHRA----- 2258
2504 WRKLPPNLSPTEIYNDGRPAKRHDIAARSHSESPSRLPINRSGTWKREHSHSSSLPRVST 2563
2259 -SKLPTKKGTA---SGSPSK-----AGSPKRIPLAPA---RRMTPORANTSLALAA 2303
2564 WRRTGSSSILSASSSESEKAKSEDEKHVNSISGTKQSKENQVSAK--GTWRKIKENEPS 2621
2304 GKSPAAASRVVSGRVSTTPPSRNSNLNGSSAAAAAKINQAQSRIANIWKRVDEAKTK 2363
2622 PTNS---TSQTVSGATNGAESKTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGN 2678
2364 QSSNLRKTQKTSNMLNANGTKPTLLR-----SSTPD-----N 2397
2679 TPPVIDSVSEK 2689
2398 TPSTAGGVKSK 2408

A:Accession: T13564
 A:Status: Preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A:Note: EG:49E4.1
 C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.4%; Score 638.5; DB 2; Length 5327;
 Best Local Similarity 18.6%; Pred. No. 6.2e-18;
 Matches 610; Conservative 504; Mismatches 1253; Indels 921; Gaps 129;

QY	2	AAASVDQLLKQVEA----	LXWNSNLRQELE--DNSNHLTKLETSANMKVILKQLQGS	55	
DB	941	ABEEIEAIAKVEAERKALLEGASAKQDSELDVPEQSKIAEVDIIATKDI	AKSR	1000	
QY	56	EDEMASSGGIDLLERLKLNLDSNPPGVKLRSKMS-----	LRSYG	97	
DB	1001	TEEQAKPAEEL-----	SSPTPEELSKTSDTKDQIGAPVDVLPVNLQESLP	1050	
QY	98	SREGSVSRSG-----	FCSVPMGSPFRPFVNGSRSTGYLBELEKERSLLIADLD	149	
DB	1051	BEKFSATIESGATTAPLPEDERIPDQIKEDLVI-----	BEKYVKEETKEAEAIVVATVQ	1106	
QY	150	-----	KEEKEKDWYVAQLNLTKRIDS-----	LPLTENFSLQTDLTRQ	188
DB	1107	TLPEAAPLAIDTILASATKAPKD--ANAELGELPDSGERVLPKMTFPAQONLLRDV	1163		
QY	189	LEY--EARQTRVAMEOLGTCQMEKKAQRRIARIQOIEKILIRIROLLOSQATEAERS	246		
DB	1164	IKTPDEVADLPVHEADLGL---YEKDSQDANAK-----	SISHKEESA	1203	
QY	247	QNHETGSHDAERNEQGVGETINMATSNGOGSTTRMDHETASVLSSTSHSAPRR---	303		
DB	1204	KEKET--DDEKENK---VGEELGDEPN-----	KVD--ISHVLLKESVQBEAKVVV	1249	
QY	304	LTSHLGTKVEMVYSLLSMLGTHDKDMSRTLLAMVSSQDSISMROSGLPLLIQLHGN	363		
DB	1250	IETTVEKKQBEIVAEATTVIQENQED---LMEQVKDKEEHOKEISGIT	1296		
QY	364	DKDSVLLGNRSGKEAPARASAAHLIHSQPDQKGRREIRVHLLEQIRAYCETCWEW	423		
DB	1297	-----	EKEAKKAST-----	PEEK-----	1310
QY	424	QEAHEPGMDQDNFMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGLQA---	IAELIQ	480	
DB	1311	ETSDITSDD--LPA---QLADPTVPKPSAKDRE-----	DTGSIESPTIEEAIE	1356	
QY	481	VDCEMYGLTNDHYSITLRRVAGMALNLPFGDVANKATLCSMKGMALVAQLKSESDL	540		
DB	1357	VE-----	VQAKQEAQ--	1366	
QY	541	QQVIASVLRNLNLRADVNSKTLREVGSVKALMECALEVK--KESTLKSVLNLSALNLSAHC	599		
DB	1367	KVPAPPEAIKTEKSPASKETSRPSATGCVKEDTEQTKSKSPVPS-----	1414		
QY	600	TENKADICAVDGAFLVGLTTRVSTNTLAIIESGGILRVNVSIIATNEHRQI---	L	656	
DB	1415	---RPSEAKDKKSPFASGEAS--RPESVAESVDEAGKASRRSREIAKTHDKDESSLDKAK	1470		
QY	657	RENNCLQTLQHLKSHSLTIVSNACGLTNLSARNPKDOEALWDMGAVSMLNLIHSKHK	716		
DB	1471	EQSRRESLAESTKPSGIDEXA-----	LASKEASRPESVTD-----	KSKPFSRRE	1517
QY	717	MIANG--SAAALRNLMANRPAYKDNIMGPSLPSLHYRKQKALFAELDAQHLSETFDN	775		
DB	1518	SIASLKAESTKDEKSAAPPK-----	EASRPGSVVESVKDETEKSKEP-----	SRRESI	1566
QY	776	IDNLSPKASHRSQRHKQSLGYGVFDTRHDDNRSDNFTGNMTVLSPYLTNTVLPGSS	835		

DB	1567	ABSAKPIEFREVSRPESVI--DGIKDSAKPESRRD-----	SP-----	LASKE	1608	
QY	836	SSR--GSLDSSRSEKORSLERERIGLGNHYPA-----	TENPGTSSKRGLOJISTAAQIA	888		
DB	1609	ASRPESVLESVKDEFIKSTEKSRRESVAESFKADTKDEKSPITSK----	DISRPESAVE	1664		
QY	889	KYMEVSAIHTSQRSSSGSTTELHCVTDERNARSSAAHTHTNTYNTFKTSENSTCS	948			
DB	1665	NVM-DAPKETSRPESAVGS-----	MKDESMKEPSPRRSRVKGDAQAQSRSTS	1710		
QY	949	MPYAKLEYKRRSNDLSNVS---SNDGYKKGQKMPKSIESYSEDEDEKFCYSYGYPADLA	1005			
DB	1711	RPASVAESAKGADDLKELSRPESTTQSKEAGSIKDEKSPLASSEARSPASVAESVKDEA	1770			
QY	1006	KHISANHWDD-----	NDGELDTPLNYSILKYSDQLNS-----	GROSFSQWERW	1049	
DB	1771	EKSKEESRRESVAESKPLPSKEASRPASVAESIKDEAKKEESRRESVAESKPLPSKEA	1830			
QY	1050	ARPKHI-----	IEDEIKQSEQ-----	RQSRNQSTTYP-----	VYTESTDDKHLKFPQHFQ	1095
DB	1831	SRPASVAESIKDEAKKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEE--SR	1889			
QY	1096	QECV-----	SPYRSRGANGSEINRVGSHNQNVSQSLCQED---	DYEDDKPTNYS	1145	
DB	1890	RESVAESKPLPSKEASRPASVAESIKDEAKKEESRRESVAESKPLPSKEASRPASVAE	1949			
QY	1146	RYSBE-EQHEEERPTNYSIKYNEEKEHVDPIDYSLKYATDIPSSQKQSFSSKSSGQ	1204			
DB	1950	SIKDEAKKEESRRESVAESKPLPSKEASRP-----	ASVAESIKDEAKSKESRR	2001		
QY	1205	SSKTEHMSGSSTSTPSSNA-----	KRONLHPSSAQSRGQKQKATCKVS-----	1252		
DB	2002	ESVAESKPLPSKEASRPASVAESIKDEAKKEESRRESVAESKPLPSKEASRPASVAES	2061			
QY	1253	-----	SINQTIOTYCVEDTPICFSRCSLSSLS--SAEDEIGCNQTTQEADS--ANTLOI	1304		
DB	2062	IKDEAKKEESRRESVAESKPLPSKEASRPASVAESKDEADKKEESRRESVAESGKA	2121			
QY	1305	ABIKGKIG-----	TRSAEDPVSEPAVQHPKTSRRLQSGSLSESEARHKAKEFPFG	1357		
DB	2122	QSIKQDOSPLKEVSRPESVAESKVDKPKSPPRRS--VAGSVTADARD--	2172			
QY	1358	AKSPKS--GACTPKSPPEHYVQETPLMFSTCTSVSSLDSPESRSIASSVQSPCSGMVSG	1416			
DB	2173	-QSPLEKSGARSPESVVDVKDEAKKEESRRES-----	KTES	2208		
QY	1417	IISPSDLPD--SPGQTMPP--SRSKT-----	PPPPPTQATQKRE-----	VPKN-KAP	1458	
DB	2209	VIPPKAKDKSPKEVLQPVSMVTETIREDAQPMKPSQAESRRRESIAESIKASSPRDEKSP	2268			
QY	1459	TAEKESGPKQAA--VNAAVQRVQVLPDADTLHLFATESTPDGFCSSLS-----	1507			
DB	2269	LASKEASRPGSVAESIKYDLDKQI--IKDKXSTHRESLEDKSAVTSEKSVSRPLSVAS	2328			
QY	1508	-----	ALSJD--	1512		
DB	2329	DHEAAVAIEDDAKSSIIPKDKSRPGFVAETVSSPIEATMEFSKIEVVEKSSLSALSIQGG	2388			
QY	1513	-----	EPFQKQVELRIM-----	PP	1527	
DB	2389	SGGKLQTDSSPVDVAEGDFSHAVASVSTVPTLTTPKPAEQIQAAGAAKTVSSPLDEALRTPS	2448			
QY	1528	VQENDNGEYTESQPKESNENQKEAKTTIDSEKDLDDDDDDDDDDDDDDDDDDDDDDDDDDDD	1587			
DB	2449	APEHISADSPAEACASEIASQDKSPQVLKESRPAPVAESKDDAAQLKSSVEDLSPVA	2508			
QY	1588	SRGKKPAQATASKLPPPPVARKPSOLPVYKL-----	LPSONRLQFQKHVSF	1632		
DB	2509	STEISRPASAGETASSPIEAPKDFAEFOAEKAVLPLTLIELKGNLPTLSSPVDVAHASV	2568			
QY	1633	TPGD-----	DMPRVYCVGPTINFSTATSLDITIEGPPNELAAGEGVRGGAQSGEPEKRD	1688		
DB	2569	QPAELSKVDIEK-----	TASSPIDEAPKSLIGSPAEPE-----	SPAESAKDAESVEKSKDA	2622	

```
QY 1699 TIP---TGRSTDEAQQ---GKTSVITIPELDDNKAEE-----GD----- 1722
Db 2623 SRPPSWESTRADSTKGDISPSPESVLEKPGDDVEKSESRPPSPVSGASITGDTKDVSR 2682
QY 1723 --ILAECINSAMPKSKHP--FRUKIMDQVQKASASSA----- 1759
Db 2693 PASVVESYKDEHDKAESRRRESIAKVESVIDEAGKSDSKSSQDQKQKSTLAKESARR 2742
QY 1760 -----PNKQLDGGKKKFTSPV---KPIQNTYRTRVRKNADSKNNLNAERFVSDNKS 1811
Db 2743 ESWVESSDKDAEKSESRPESVIASGEVPR-----ESKSP-----DSKDT 2783
QY 1812 KQONLKNNSKDFNDKLPNNEDVRGSPAFDPSHHVTPTEGTPYCFSRNDSLSSLDFFDDDD 1871
Db 2784 SRPGSVESVTADE--KSEQQSRRESVAESVKADTKDKGSKQASRPSSVDELLKDDDE 2841
QY 1872 VDLRS-----EKAELKAKENKESSEAKVTSHTLTNSQOS--ANK 1909
Db 2842 KOESRRQITGSHKAMTWGDESPMDKADKSESPSPESVAESIKH-ENTKDESPLSGR 2900
QY 1910 TQAIH---KQPINRQOPKPIIQKQSTFPQSSKDPDRGAATDEKLQHPA----- 1955
Db 2901 RDSVAESIKSDITKGEKSPLESKEVSRPESVY---GSIKDEKAESRRRESVAESVSPES 2955
QY 1956 --IENTPVCFSHN---SSLSGLSD-----IDQENNKENE- 1985
Db 2956 SKDATSAPPSKEHSPPEVSLGDKDGKTSRRVSVADSIKDEKSLVSGEASRPSEEA 3015
QY 1986 -----PIKETEPDPSQGPSPKQASGYAPKSPHVEDTPVCFSPRNSLSLSIDSED- 2036
Db 3016 ESKLDAAPSOETSPESVTSVDGKSPVASKE-----ASRPASVAENAKDSADE 3066
QY 2037 -----DILQBCISAMPKKPSLKGDNKHSPPNMGWILGEDTLTLDLDIOPPOS-E 2089
Db 3067 SKEQPESLPQSGISIKDEKSPKASKDEAKSEESRRRESVAEQPPLVSVKEVRPASVA 3126
QY 2090 HGLSPDSDFWKAIQEGANGSVSSLHQAACLSROASDSDSILSLKSGISLGSPPH 2149
Db 3127 ESVKDEAE---KSKEE-----SPLMSKEASRPASVAGSVK----- 3158
QY 2150 LTPDDEERFPTSNKGPRIKDPGEKSTLTTKIESKGIKGGKVKYKSLITGKVRNSEI 2209
Db 3159 ---DEAEK---SKEESRRRESVAESKPLPSKE---ASRPASVAESVKQPADKSEESRR- 3208
QY 2210 SGOMKOPL---CANMPSISRGRTM-HIPVNRSSSSSTPVSCKGPPKLTTPASKGPSEGT 2266
Db 3209 SGAESKPLASKEASRPA-----SVAESIKDEAKSEESRRRESVAESKPLPSKE 3257
QY 2267 ATTSRGAKPSVKSELSPVAKOTSQIGSSKAP--SRSGSRDSTPSRPAQPLSRPIOSP 2324
Db 3258 A-SRPTSVAESVKDEAKSEESSRDSVAESKPLASKEASRPASVAESVQDEAKSEES 3316
QY 2325 GRNSISPGRNGTSPNKLSQLPRTSSPSTASTKSSGSGKWSYTSFGROMSQONLTQOTGL 2384
Db 3317 RRESVAE---KSLPAYKEASRPASVAESIKDEAKSK-----ESRRRESVAESKPL 3364
QY 2385 SKNASSIPRESASKGLNQMNNGANKKVELSRMSSTKSGSESDRSERPVLVQSTFI 2444
Db 3365 ASKEASRPTSVAESV-----KDEAKSEESSRDSVAESKPLASKEASRPASVAES--V 3416
QY 2445 KEAPSTLRKLEESAPES---LSPSRPASPPTSQAQTPVLSPLDMSLSHTSSVQA 2501
Db 3417 QDEAKSEESRRRESVAESKPLASKEASRPASVAESVKD-----DAKSEES--- 3464
QY 2502 GQWRKLPPNLSPTIEYNDGRPAK----RHDIAHSHSESPSLRPINRSGTWKREHSHSS 2556
Db 3465 --RRRESVAESKPLASKEASRPASVAESVKDEAKSEES-----RRRESVAESK 3510
QY 2557 SLPRVSTWRTGSSSILSASSESEKAKSEDEKHVNSIGTKQSKENQV--SAGTWRK 2614
Db 3511 PLPSKEASRPTSVAESVKDEAKSEESRRRESVAESKPLASKEASRPASVAESVKDEAK 3570
```

```
QY 2615 IKEN-----EPSPINSTSQTVSSGATNGAASKTLIYQVAFAPVSKTEDVWVRIEDCPIN 2667
Db 3571 SKEESRRRESVAESKPLASKE---ASRPASVAESVKDEAKSEKVESRRSV---AEKSPLP 3624
QY 2668 NPBSGRSPGTGNTPPVDSYSEKANPNIKSKDNQAKNQVNGSVPMRTVYGLNRLTSTFQ 2727
Db 3625 SKASR-PTS---VAESVKDEADKSEER-----RESGAESKPLASME 3664
QY 2728 VDAPOKGTETKPGQNNPVPVSTNETSPIVERTPPSSSSSSSKHSPSGTVAARVTFNPN 2787
Db 3665 ASRPTSVAESVK---DETEKSEESRRRESVTEKSPLESKEASRPTSVAESVKDEAK-SKE 3721
QY 2788 PSRPSKASADST-----SARPSOIPTPVNNNTK--KXDSKTDSTESSG 2827
Db 3722 ESRRESVAESKPLASKESSRPASVAESIKDEAGTYKQESRRRESMPESG 3769

RESULT 6
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #ext_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Ima, A.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BA043752.1; GSPDB:GNC0149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

Query Match 3.6%; Score 518; DB 2; Length 2271;
Best Local Similarity 19.5%; Pred. No. 1.8e-13;
Matches 452; Conservative 333; Mismatches 1000; Indels 534; Gaps 76;

QY 606 ICADVGMALFLVGLTYRSTNTLAIE-----SGGIL-----RVSSLI 646
Db 353 IGLSNAGFKPLDTYHNTSPNSAAKANADPSNAGGAFVTTDSYGVATYTTSSST 412
QY 647 ATNEDHQILRENNCLQTLQHLKSHSLTIVSNACGTLNLSARNPKDQEAALMDGAVSM 706
Db 413 ADNAAKLVQPTNTTQDFDINVGDTKMTVYQAQTW---TNTISD---W---IAK 461
QY 707 LKMLHSHKHVIAWGAAALRNLM-----ANRPAYKDA-----NIMSPGSSLP SLH 753
Db 462 SGTNFSLSMTASTGGATNLQVQFGTFEYTESAVTQVRVYDVTGKDIIPPKTYSNVD 521
QY 754 V-----RKOKALEAELDAQHLSETFNIDNLSPKASHRSKORHKQSLYGDYVDTNRHDD 808
Db 522 QVTTIDNQOQSALTAK-----GNYTISVDSYSYASTYNDINKTVKMTNAGQSV--TYFTFD 573
QY 809 NRSDFNFTGNMTV-LSPYLNTTVLPSSSSSRGSLDSRSRSEKRSLERRIGIGL-NYHPA 866
Db 574 VKAPTIVVGNQTEVGVGTMPVPLTTTDDNGTGTVNTVT-----GLPSGLSYDSA 623
QY 867 TEN-PGTSRRKRGLOISTTAAQIAKVMVEVAIHTSQEDRSSGSTTLHLCHVDTNRNALR-- 923
Db 624 TNSIIGTPTTIG-----QSTVTVTVDQANNKSTTTFTINVVDVTAPTPTPI 670
QY 924 --RSSAAHTSNVTNFTKSENSNRTCMPYAKLEYKRSNDNLSNDSNDSDYGRKGOMKP 981
Db 671 GDQSSVYSPISPIKATQDNSGNVNTVTGLPSGLTFDSTNNTISGT----- 719
QY 982 SIESYEDDSKFCQSYQYPADLAHKIHSANHHDDNDGELDTPIINYSL----KYSDEOLNS 1038
Db 720 -----PTNIGTSTISIVSTDASGNKTTTTFKYEIVTRNSMDSYSTS 760
```


Query Match		3.5%; Score 517; DB 2; Length 2232;
Best Local Similarity		20.4%; Pred. No. 2e-13;
Matches 387; Conservative 247; Mismatches 811; Indels 454; Gaps 69;		
QY	1185	TDIPSSQKQSFs-----FKSSSGGSSKTEHMSSESTSTPSSNAKRONQLHPSSAQSRs 1240
Db	360	TVVPGSSSTFASSTPIASSSPGSTVTVAPGSSSTYGSSSTPSSAGSSSGTMTSGSTGS 419
QY	1241	GOPOKAATCKVSSINQETIQYCVEDTPICSRc--SSLSSLSAEDIEGNCQTQBADs 1298
Db	420	-----TVVAVSSSTFGs-----STPIASSSSSGSTTVVVGSSSTYGSTSTPSSSSs 468
QY	1299	ANTLQIAIKIGKIGTRSAEDPVSEPAVSQHPRTKSSRLQg--SSLSSSARHRAVEFPs 1356
Db	469	AGT--ASTISGSGTATVPGSS--SSVSGSTQASPSPGMTSVSGTGTSTTV-VPG 524
QY	1357	CAKSPKSGAQTPKPPPHYQETPLMFSCRc-----TVSSLLDFESRSIASSVQSE 1408
Db	525	SSTSPAPSSSNPSSPASTGSTITIGSSSIIIVTSGSTVSGTSTGTSQTSLASSTATP 584
QY	1409	PCSGMVSGIIGSPDLDPDP-----GOTMPPSRSKTTPPPPPQTAQKREVPKKAPTAKRE 1464
Db	585	GSSSTVPSSSPQPSQAPANTGSTITPSQTSQSPSPSMN-----PSSSTTIGSSQS 637
QY	1465	-----SGPKQa-----AVNAAVQVQVLPDADTLHLHFATSTPDGFCSSSLGALS 1510
Db	638	TITPEGSTASGPTGSGTSTFVATEVTQSTVPSSSLGTQSTNSPSSPSSLPSTSGMS 697
QY	1511	--LDPPFIQKQVELIAMPVQENDNGNE---TESEQPKESNENOEKEAE----- 1554
Db	698	TLTSEP-----SPSSTQSSGAQSTTTTPSPNSQSTSLSESTGATTSAGSAG 746
QY	1555	--KTIDSEKLLDDDDDDIEILEBECIISAMPTKSRKGRKPAQPAKLPVPVARKPSQL 1612
Db	747	TTMTSPSSSSVGGSSQG-----STPAASTTSGEMTSQGSTQTPGSSSVSTSAAI 795
QY	1613	PVYKLLPQNRLQPKHVSFTPCDDMPRVYCEGTPINFSTAT-----SLSDLTIESPP 1666
Db	796	-----LTJTCQSVSTN-----SPGSTVTRPSTVSGSTVSGSTVTVGSTEASTSGSSVASS 846
QY	1667	NELAAEGVRGGAQSGE-----FEKEDTITPEGRST-DEAQQGKTSVVIPE----- 1712
Db	847	PAPSTSONPSTSSGSSMITQSPYPSQSTSPVESSTTPSPGSPGTLTSTSPSQSTT 906
QY	1713	LDNKAERGDILAEICINAMPKSHKDFRVKKIMDQVQOASASSAPNKNQLOGKKKP 1772
Db	907	IGSTQGSTSPGISTTSEEMTSQGSTQTPGSTVTPQSTVSDSTSGSTVTV-GSTEGS 965
QY	1773	TSPVKPIPONTBYRVRKNADSKNLNAERVFSDNKKONLKNKSKDFNDKLNNED 1832
Db	966	SPIPSTQNTNPST-----SSGSSMTQTPQSSQSTSPVESSTSGA----- 1007
QY	1833	RVRGSFAPDSPHHYTPTEGTPYCFGRNDSLSLDFDDDDVLSREKAE LRKAKENKESEA 1892
Db	1008	-----TSSSGSP----- 1014
QY	1893	KVTSHELTSNCOOSANKTQAIKQPINRGOPKPILOKOSTPPOSSKDIPTDRGAATDEKLO 1952
Db	1015	-----GTLTISIPSPSPSTIGs-----SOGSTSPV---STISQGSTETPGSTGTVK-- 1062
QY	1953	NPAIENTPVCFSHNSLSLSLDIOE-----NNKENEPIKETETPPDQCEPSPKQAS 2005
Db	1063	-----PSTVSGSASSGSTATMGTEASSTSGSSSTPNPFSQSTSPSTs-GATSSPGSS 1114
QY	2006	GYAPKSFH---VEDTPVCFERNSSLSLSIDSEDLLECISSAMPKKKPSRLKGDNEK 2062
Db	1115	GTLTISIPSPSQSTIGSSQSTSPVSTSTSGDMTSQG---STQTPGSTGTSTVTPQSTGS 1172
QY	2063	HSPRNNGGTLGLDLTLDLKIDIRPDSEHGLSPDSENFOWKAIQECa-----NSIVSS 2114
Db	1173	GSTISGEITSQGST-----QTPRSSLSTSPASTSTQQQSVSTNSPGSTVTPQSTVRGS 1226

RESULT 8
23327
adenomatus polyposis coli protein 1 - Caenorhabditis elegans
N:Alternate names: apr-1 protein
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23327; T03822
R:Gardner, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19727
A:Accession: T23327
A>Status: preliminary; translated from GE/EMBL/DBU

QY	2115	LHQAAAAACLSROASSDSIILSKSGISLGSPFHLT-----PDQEEK 2157
Db	1227	TSSGSIVTTGSTECSSTSGS--SSATSLSSSPVPSTSQSPNPSTSGSTPTPNPQSSTs 1284
QY	2158	PFTSNK-----GPRILKP-----GEKSTLETKKIESKGIKGG----- 2191
Db	1285	PVYSTTTGEMTHSGSTQTPTSTIGSTVTPSTVSGNSSGSTVTIGSSEASTSGSFKTSP 1344
QY	2192	-----KKVYKSLITGKVRN-SEISGOMKQPLQANMPISIRGRTHIHPGVR 2237
Db	1345	SSISPVPTSPSPITTFASSTSGSTISDVSSVSTTSLAFLSSSLPT-----VPSST 1396
QY	2238	NSSSSTSPVSKGPPPLKTPAKSPSPSGQTATTTPRGAKPVSXKSELSPVARTQTSIQGSSK 2297
Db	1397	QSPSTSESSSK-----ASSSPVPSQTSTP---TNPTGSTESSTLL--SSTISGSTQ 1444
QY	2298	--APSSGSRDSTPSRPAQPLSRPIQSPGRNIS--PGRNGISPPNKLQ-----LP 2346
Db	1445	HTMWSKASGSGSTSPSTNSQTGSTVTVMGSSSTSGVSTSSASSTQPMSTQCGSAGSTVAS 1504
QY	2347	RTSPSTASTKSSGSKMYSTFGQMS--QQNLTKQTGLSKNASSIPRSESASKGLN-- 2402
Db	1505	STASPAASSTAPSTGTMTSGSTSGTVGSTISESTTASASSQTGSTVTVMGSSSTSGVSTs 1564
QY	2403	-----QM--NMGANKKVELSRM-----SSTKSSGESDRSERPVLVRQSTFI 2444
Db	1565	SASSTQPMSTQCGSAGSTVASSTAGLVSTSTVPSSTGTMTGTSSTSGTVGSTISEST-T 1623
QY	2445	KEAPSTPLRKLKEESAFESLSPSRPAPSTRQ-----AQTPLVSPSLP 2489
Db	1624	ASASSQTGSTVTVMGSSSTSGVSTSS--ASSTQPMSTQCGSSAGSTVASSTGLVSTSTV 1681
QY	2490	DMSLTHSSVQAGWRKLPNLSPTTIEYDGRPAKHDIARSHSESPRLPPINRSGTWKR 2549
Db	1682	PSSTGTMGSTSSG-----TVGSTI-----SESSTAASASSQTGSTVTVMGSSSTSGV 1727
QY	2550	EHSKHSSLPVSTWRTTSGSSSILS--ASSESEKAKGEDEKHVANSISGTKOSKENQVS 2607
Db	1728	STASASGQPMSTQCGSAGSTVSTAGTAPASSTAPSTGTMTSGSTSGTVGSTVSTQSS 1787
QY	2608	AKGTWRKIKENEPSTNSTSQTVSSG-----ATNGAERKTLI 2644
Db	1788	TAA-----STTSHTGSTVTLGSSSTSSNQMGTSQGSVSGSTVASSTAGLVSTSTV 1837
QY	2645	YQAPAVSKTEDVWVRIEDCPINNPSGRSPTGNTTFFVIDSVSEKANPNIKDSKDNQAKQ 2704
Db	1838	PSSTGTMGSTSGTVGSTISESTTASASSQTGST--VTMGSSSTSGVSTSSASSTQPM 1895
QY	2705	NVNGNS-----VPMRTVGLNRLTSPITQVDAPQKGTETKPGQNNPVVSET-NESPIVE 2758
Db	1896	STQGSAGSTVASSTAGLVSTST-----VPSSTGT---MGSTSGTVGSTISESTAA 1946
QY	2759	RTFPSSSSSKHSPSGTVAa---RVTPFNYPSPRKSSADSTARSQIPTPVNNTKK 2815
Db	1947	STSQSTGTSTIGSTGCTNPSRSLSQITITPSPQSTESTQTSLPSSSPSPSTHVSs 2006
QY	2816	RDSKTDST-----ESSGTQSPKRHSGLVTS 2842
Db	2007	SEGTMSSGATTSQDKMSFLSSTGTTVSSSRGSSLAT 2045

A:Molecule type: DNA
A:Residues: 1-1186 <WIL>
A:Cross-references: EMBL:Z75712; PIDN:CA00045.1; GSPDB:GN00019; CESP:K04G2.8a
A:Experimental source: clone K04G2
R:Rocheleau, C.E.; Downs, W.D.; Lin, R.; Wittmann, C.; Bei, Y.; Cha, Y.H.; Ali, M.; Prie
Cell 90, 707-716, 1997
A:Title: Wnt signaling and an APC related gene specify endoderm in early C. elegans embr
A:Reference number: Z15051; MUID:97433081; PMID:9288750
A:Accession: T03822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1186 <ROC>
A:Cross-references: EMBL:AF013950; NID:G2338717; PIDN:AAC47747.1; PID:G2338718
A:Experimental source: strain Bristol
C:Genetics:
A:Gene: apr-1; CESP:K04G2.8a
A:Map position: 1
A:Introns: 61/3; 113/2; 355/3; 551/3; 1067/3; 1161/3
A>Note: apr-1

Query Match 3.5%; Score 516.5; DB 2; Length 1186;
Best Local Similarity 21.9%; Pred. No. 9.1e-14;
Matches 289; Conservative 216; Mismatches 562; Indels 251; Gaps 52;

Qy 271 MATSGNGGSTRMDHETASVLSSTHSAAPRLTSHLGTKEVYVLSMLGTHDKDM 330
Db 1 MSSSSDENETT--IHRGTGNTGGSGIYQPRAGSSKRTSNVRHDVS-----DVIDE 50

Qy 331 SRTLAMSSQDSCISMRQGCPLLIQLLHGNKDVSLLGNRSGSKARASAAAHN- 389
Db 51 EEHVARP--REDAIEV--DDAIVLSSLSLHFHKRDIVPTDEDNK-----LRELHEK 100

Qy 390 ---IHSQPDKRGREIRVHLLEQIRAYCETCWEQEAHEFGMDQKPNPAPVEHQI 446
Db 101 IFALITSESQVNRKRLKALPASNCVR-----EQVY--LRKPKSTPPASYHEL 149

Qy 447 CPACVLMKLSFDEEHRHANNEGLGGLQAIELQVDCMYGL--TNDHYSITLIRYAGNA 504
Db 150 NAALHTIVKESFEEYRKAVTVLGLVEALAEVILEVHTFGINETPGEHRNIRKLIANA 209

Qy 505 LTNLTFDGVANKATLCMKGMCRALVAQLKSESEDQOVIASVLRNLSRADVNSKRTLR 564
Db 210 LTNLTYQIHSKRRLCSYDGFRCVW--RVIESFNITQVYAGLIRNLSNADSGMSEALQ 268

Qy 565 EVGSVKALMECALEVKES--TLKSVLSALWNLNLSAHTENKADICAVDGAFLVGLTLYR 623
Db 269 P--TVHALSTAAVHAHTRFDVTATLSALWNLNLSAGHVENKRTICDTPNCLKVLASLLSP 326

Qy 624 SQNTLAIISGGIIRNVSLIATNEDHQILRENNCLQTLHLKSHSLTIVSNACGT 683
Db 327 ARTSL--VDSATGILKYVQYLANSTHLEL--RSLLITRMLTLKLSASFCTVNTLGA 382

Qy 684 LNLISARNPKDQAL--WDGAVSMKLNLIHSHKMTIAMSAAALRNLMANRPAYKQANI 742
Db 383 IANLIVKDPHMQMIRODMAAQQVLANVLRNSNRDDIRTAVKSVLNTL--NQPCSHRYGDM 440

Qy 743 M-SFGSSLSLHVKKALAEADAHLSETFNIDNLSFKASHRSKQRHKSGLYGDYVF 801
Db 441 SHSVGGATGCMQLSEPOLQVQ--TSHAYHGTASPRLLSLRATRASPGKYIQGAQQQLI 499

Qy 802 DTRHDDNRSD---NFTNGMNVLSPLYNTVLPSSSSRSGSLDSRSBK--DRSLEREG 857
Db 500 QTPQVDQRSSSLPRHF-----AVORNGFVMAQSVNQ--MDQHQQQQMIYQLQOQQQ 549

Qy 858 IGLGNVHPATENPCTSSKRGQLISTTAAQIAKYMEEVSALHTSOEDRSSGSTTELHCV-- 915
Db 550 I-----MTEDQAQWEHH-----QQIMYLQOQQQFQHQIQOQQQMKQAQADPVP 594

Qy 916 TDERNALRRSSAAHTSNTYFTKSENSNRCTSMYPKLEYKRSSNDLSNYSNDGYGK 975
Db 595 TDDLDIPTVMGTRSNRSRSLGSMNPGSVMTNMNSSLDTAANSRSLSPSYND--IPA 653

Qy 976 RGQMKPSIESYSEDDSKFCSCYQYFADLAHKTH-----SANHMDNDNGELDPIN----- 1026

Db 654 SPTMCAQVFNLPKSTESE---HHQLTSQOQNTTHYSSGSANTMTSRDGAITVPMDNIITP 710
Qy 1027 -YSLK---YSDEQNSGRQSPQNERWARPKHITIDEIKQSEQ-----RQSRNQ 1072
Db 711 TYAILNPILVHEQTPNGTVPKTSSELDSPDVLPGPSLEBEGDYAIIGAAOKTODEL 770
Qy 1073 TTPVYTE---STDDKHLKFPHEG-----QCECVSPYRSRGANGSETNRV--GSNHGI 1121
Db 771 LTRSIQSEMPSTSTPKKVPKVSPLRNGFFSTQKTTTSPAWSH-----PDTSP 826
Qy 1122 NONVSQSLCQDDY-----EDDKPTNYSERYSEEBEHEEPTNYSYKYNEEKHVD 1174
Db 827 QPNRRQDASDADRLMESIMSEMPKRIISPRLAGTQCYLEPE--PERRSHSKNEADRRD 885
Qy 1175 -----QPIDYS-LKYATDIPSSQKQSFSSKSSGSSKTEHMSSESSENTSTPSSNAKR 1227
Db 886 AFTASHEPDSHNGIDVARGSDWSPQQQLHMESELESQASSEDSFGLTAEFPNGSTGA-- 943
Qy 1228 QNLHPSSAQRSGQPOKAATCKVSSINQETIQTQYCYVEDTPICFSS----- 1272
Db 944 -----AANTMRFDDEIDASLPWCDDDDYDTHDFEDYEDEEDPDPA 986
Qy 1273 -----RCSLSLSLSAEDEIGCNOTTOEADSANTLOIAELKGIKIGTRSAEDP 1319
Db 987 TQFDGVDQAQITDCSMITSGSSQR---NETTTSRDSKALATSTPKGS----- 1034
Qy 1320 VSEVPAYSQHPRTKSRRLQSSLSSESARHKAVFFPSGA---KSPSK-SGAQTPKSP-- 1373
Db 1035 ASSLPGVQATR-----VSTNGKSRPLVPKTNGLSLVDKPKPIIASRRPLPKP 1084
Qy 1374 -----EHVQTPLMFRCTSVSLDSFERS-----IASSVQSPSCGMVSGIISPS 1421
Db 1085 TLLKDKHYPEE-----DSIENQTRDDTIYNVAPVVEAQERIMYKALKQOK 1130
Qy 1422 DLPDSP--GQTMPSPRSKTPPP-----PPTAQTKREVPKNKAPTAEKRESGPKQAAV 1472
Db 1131 NIEQSPSICNGSPIAKSAIVTPYNYQKPTPTGRNNGEMSKSVT-----PNPKQMLV 1183

RESULT 9
T23330
hypoetical protein K04G2.8b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23330
R:Gardner, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: 219727
A:Accession: T23330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1188 <WIL>
A:Cross-references: EMBL:Z75712; PIDN:CA00048.1; GSPDB:GN00019; CESP:K04G2.8b
A:Experimental source: clone K04G2
C:Genetics:
A:Gene: CESP:K04G2.8b
A:Map position: 1
A:Introns: 61/3; 113/2; 355/3; 551/3; 1069/3; 1163/3

Query Match 3.5%; Score 514.5; DB 2; Length 1188;
Best Local Similarity 21.9%; Pred. No. 1.1e-13;
Matches 289; Conservative 216; Mismatches 564; Indels 249; Gaps 52;

Qy 271 MATSGNGGSTRMDHETASVLSSTHSAAPRLTSHLGTKEVYVLSMLGTHDKDM 330
Db 1 MSSSSDENETT--IHRGTGNTGGSGIYQPRAGSSKRTSNVRHDVS-----DVIDE 50

Qy 331 SRTLAMSSQDSCISMRQGCPLLIQLLHGNKDVSLLGNRSGSKARASAAAHN- 389
Db 51 EEHVARP--REDAIEV--DDAIVLSSLSLHFHKRDIVPTDEDNK-----LRELHEK 100

Qy 390 ---IHSQPDKRGREIRVHLLEQIRAYCETCWEQEAHEFGMDQKPNPAPVEHQI 446
Db 101 IFALITSESQVNRKRLKALPASNCVR-----EQVY--LRKPKSTPPASYHEL 149

Qy 447 CPACVLMKLSFDEEHRHANNEGLGGLQAIELQVDCMYGL--TNDHYSITLIRYAGNA 504
Db 150 NAALHTIVKESFEEYRKAVTVLGLVEALAEVILEVHTFGINETPGEHRNIRKLIANA 209

Qy 505 LTNLTFDGVANKATLCMKGMCRALVAQLKSESEDQOVIASVLRNLSRADVNSKRTLR 564
Db 210 LTNLTYQIHSKRRLCSYDGFRCVW--RVIESFNITQVYAGLIRNLSNADSGMSEALQ 268

Qy 565 EVGSVKALMECALEVKES--TLKSVLSALWNLNLSAHTENKADICAVDGAFLVGLTLYR 623
Db 269 P--TVHALSTAAVHAHTRFDVTATLSALWNLNLSAGHVENKRTICDTPNCLKVLASLLSP 326

Qy 624 SQNTLAIISGGIIRNVSLIATNEDHQILRENNCLQTLHLKSHSLTIVSNACGT 683
Db 327 ARTSL--VDSATGILKYVQYLANSTHLEL--RSLLITRMLTLKLSASFCTVNTLGA 382

Qy 684 LNLISARNPKDQAL--WDGAVSMKLNLIHSHKMTIAMSAAALRNLMANRPAYKQANI 742
Db 383 IANLIVKDPHMQMIRODMAAQQVLANVLRNSNRDDIRTAVKSVLNTL--NQPCSHRYGDM 440

Qy 743 M-SFGSSLSLHVKKALAEADAHLSETFNIDNLSFKASHRSKQRHKSGLYGDYVF 801
Db 441 SHSVGGATGCMQLSEPOLQVQ--TSHAYHGTASPRLLSLRATRASPGKYIQGAQQQLI 499

Qy 802 DTRHDDNRSD---NFTNGMNVLSPLYNTVLPSSSSRSGSLDSRSBK--DRSLEREG 857
Db 500 QTPQVDQRSSSLPRHF-----AVORNGFVMAQSVNQ--MDQHQQQQMIYQLQOQQQ 549

Qy 858 IGLGNVHPATENPCTSSKRGQLISTTAAQIAKYMEEVSALHTSOEDRSSGSTTELHCV-- 915
Db 550 I-----MTEDQAQWEHH-----QQIMYLQOQQQFQHQIQOQQQMKQAQADPVP 594

Qy 916 TDERNALRRSSAAHTSNTYFTKSENSNRCTSMYPKLEYKRSSNDLSNYSNDGYGK 975
Db 595 TDDLDIPTVMGTRSNRSRSLGSMNPGSVMTNMNSSLDTAANSRSLSPSYND--IPA 653

Qy 976 RGQMKPSIESYSEDDSKFCSCYQYFADLAHKTH-----SANHMDNDNGELDPIN----- 1026

Db 442 DAKKDEKRTKPBVKXLSKPDLPKF-----TPEVRKTLXKAKAPGRVKVDKGRA 491
QY 1029 LKYSDQNLNSGRQSPSQNE-----RWARPKHIED--EIKQSEQRQSNQST 1073
Db 492 ARGEKUSSEBTPPAQGAAPAAVSGHRELALSSPEDLTQDFEELKREERGLLAEQRD 551
QY 1074 T-----YPVYTESTDDKHLK-----FQP-----HFGQECVSPYRSRGANGSETNR-- 1114
Db 552 TGLGEKPLPADATQOQHPSAAIQTQSGPVLGEHVEREKEVVP--DSPDKGS--TNRGP 609
QY 1115 -----VGSNHNQNVQSQCQEDDDYDDKPTNYSRYSEBQ 1152
Db 610 DSGAEVEKEKETWBERKQRAELGPENTAAARESEAEVKEDVIEKAELEEMETHPSDEE 669
QY 1153 HEEBERPTNYSIKYNEEKRHVDOPIDYSLKYATDI-----PSSQKQSPFSKSSG 1203
Db 670 GEETKAESFY-----QKHTQALKAAPKSRALGGRDLGFGKAPKEKETAFLSSLAT 722
QY 1204 QSSKTEHMSSENSTPSSNAKR-----ONQLH-----PSSAQSRSQ-----PQ 1244
Db 723 PAGATEHVSYIQDETIPGYSETEQISDEBIHDEPDERPAPPFRPTSTYDLSGEGPGPF 782
QY 1245 KAATCKVSSNQETIQYCYVEDTPICFSRCSLSLSLSSAEDEIGCQNTTQEADSAANTLOI 1304
Db 783 EASQAADSAPASSSKTYGAPETELTY--PNVVAAPLAE--HVSSATSITE 832
QY 1305 AEIKGKIGTRSAED--PVSEVPVAVSOHPRTKSSRLQSSLSSESARHKAVE-----FPS-CA 1358
Db 833 CDKLSSPATSVABDQSVASLTAQTEETGKSLLDVTISPSRTEATQGLDYVPSACT 892
QY 1359 KSPSKSAQTP--KSPPEHVQVETPLMFRCSTSVSLDSFE-----SRSIASS----- 1404
Db 893 ISPTSLSEEDKGFKSP-----CEDFSVTGESEKKGTVGRGLGEKAVGKE 939
QY 1405 -----VOSEPCSGMVSGIISPSDLPDPGOTMPPS-----RSKTPPP 1441
Db 940 EKVVTSEKLSGQYAAVFG-----APGHTLPGEALGEVEBERCLSPDDSTVKAASPPP 993
QY 1442 --PPQTAQTVREVPKNAKPAEKR-----RVQVLPDADTLHFAFESTPDGFCSSLSA 1508
Db 994 SGPPSAHA--PFFHQSPVEDKSEPRDFQEDSWGETHKSPGVSKEDSBEQTVKPGPEG 1049
QY 1471 -----AVNAAVQ-----RVQVLPDADTLHFAFESTPDGFCSSLSA 1508
Db 1050 TSEBGKPPTRSOAQDMPVSIAGGTGCTIQLLPEQDAIVFETG-----EAGSNLGA 1103
QY 1509 LSL-----DBPFIQKVELR-----IMPPYQENDNGNETESEQP 1542
Db 1104 GTLPGEVRTSTEBATEP--QKDEVLRAFTQSLSPEDAELSLSVLVSPDTTQEAATPRSP 1161
QY 1543 KESNENQKEAEKTIIDSEKDL-----LDDSDDDIEILERCIIISAMPTKSRKKGKPAQTAS 1599
Db 1162 CSLKEQO-----PHKOLWMPSPEDTQSLSFSE--SPSKET-----SLDISS 1202
QY 1600 KLPPPPVA-----RKPSOLPVYKL-----LPSQNRLOPKHVSPTPGDD 1637
Db 1203 KQLSPESLGTQFQELNKEGPGVWKAEDDSCHLAPVSIPEPHRAIVSPSTDTTPTAGT 1262
QY 1638 MP-----RVYC-----VEGTPINFSTATSILDTIESPPNELAA-----GE 1673
Db 1263 LPQGSFHSBALSVDRKHSPEITGPGHFMT-----SDSSLTKSPESLSPSAMEDLAVWE 1318
QY 1674 GVRGAQSGEFKRD-----IP-----TEGRSTDEAGGKTSVTIPE 1712
Db 1319 GKAPGEKEPELSETROQKQILPEKAVVRODIIHQDQALDEENKPGRQDXT--PE 1377
QY 1713 LDNKAEBGDIILACINSAMPKGGKHPRVKIIVQVQOASASSAPNKGDLQKKXP 1772
Db 1378 QKGRDLDEKTAALDKGPEPK-----EKDLREDQGORAGPPAEKOKASQORDTD 1428
QY 1773 TSPVFKPIQNTVTRKRVKADSKNLNAERYFSNKKQNLKNKSKDFNDKLPNNED 1832
Db 1429 LQOT-----QATEPRDAQERR-----SEKDKSLELDRPTPEKDRILVQED 1472

QY 1833 RVRGSAFDSPHHYTPIEGTYPYCFSRNDSLSLDFDDDDVLSREKAEALRKAKE---NKE 1889
Db 1473 R-----APEHSIP-----EPTOTDRAPDRKGTDDKEQKEASEBEKEQVLEKQDWALKE 1521
QY 1890 SEAKVTSHELTSNCOASANKQTAIAKQPIKNGQPKPILOKOSTPQSSKOIPDRGAA--- 1946
Db 1522 GET-----LDQEARIAEKDELKEDTKQOQKSSFFVEDKTT--TSKETVLDQKSAKA 1572
QY 1947 --TDEKLQNFJAIENTPVCFSHNSSLSLSDIDQENN--KENEPI---KETEPPDSQGBP-- 1999
Db 1573 DSVEQDGALEKTRALGLEBSPAEGSKARQEKYKWEQDVVQWRETS--TRGEPVG 1630
QY 2000 ---SKPQASVAPKS--FHVEDTPVCFSRNSSLSLSDISEDDLLQSCI--SSAMPK--K 2050
Db 1631 CQKEPVPWEGKSPQEVRVYWRDRDITLQDAYWRELSCDRKVWFHELQGGQARPRYCE 1690
QY 2051 KKPRL--KGDNE-----KHSPRNMGILGEDTLDLKDQIRPDSEHGLSPQSENEDWK 2102
Db 1691 ERESTFLDEGDEQEIITPLQHTPS-----PWTSDFKDFQBPPLKQGLEVER---WL 1739
QY 2103 AIOEGANSIVSSLIHQAAAAACLSQASSDSLSLSKSGISLSPFHLTPDOEEK-----P 2158
Db 1740 A-----ESPVLPEEEDKLTTRSP 1758
QY 2159 F-----TSNKGPRILKPEKS-----TLETKKIESEK----- 2186
Db 1759 FEIISPPASPEMTGORVPS--APQOESPVDPTSTAPMNEPTTPSWLAIEIPWPVKDR 1816
QY 2187 -----GIKGGKVKYKSLITGKVRNSEISGOMQOPLQANMP 2222
Db 1817 PLPPAPLSPAPPTPAPEPHTPVFWSGLAEYDVVAQVQEALEGGPYSPGLKDYR 1876
QY 2223 SISRGRMIHPIGVNRNSSSTSPVSKKGPLKT-----PASKSPSE----- 2263
Db 1877 KAEGERECEGAGAPDSSSPKVPKPEAGESLATRDTQTEPEQREPTPYPDERSFOVADI 1936
QY 2264 -----QQTATT--SPRGAK-----PSYKSELSPVARQTSIGGSKAPSPSGSRDST 2308
Db 1937 YEQMLLATGLGPACTREPPPLGASGDWPHLSTKEAAGCNTSAKETSSPASPQLQSDT 1996
QY 2309 P-----SRPAQOPLSRPIQSGRN--SISFG-----RNGISPPNKLSQLPRT 2348
Db 1997 PAFSYASLAGPVPPOEP--DFGNVPEPITPPVPPRAPISLSDKDLSPPLNGSTVSCS 2054
QY 2349 SSPSTASTKSGSGKMSYTSFGRQMSQNLTKOTGLSKNASSIPRSESASKGLNQ---NN 2405
Db 2055 PDRTPSPKETGRHW-----DDGTNDSDLEKGAQEQPEKETSPPSPHHPMPWG 2103
QY 2406 NNGANKKVELSRMSSTKSSGS----- 2427
Db 2104 HSLWPETEAYSLSDSHLGSVRLDPFASFGFSSLLQAPPLPQSPAPERSAPCGSL 2163
QY 2428 --ESDRS-----ERFVLVROSTFIKEAPSPTLRKLKEEGASPFES---LSPSSRRASP- 2474
Db 2164 AFGCDRALALVPGTPTFTRHDEYLEVTKAPSLDSSLPOLPSPSSPGPILLSNLPASFA 2223
QY 2475 ---TRSOATQTVLSPSL-----PDMSLSTHSSVOAGG-----WRKLPPNLSPTIY 2517
Db 2224 LSEGSSEATTPIVSSVAERFPFGLBAEQSARGLSGKESAAHSLWDLTP--LSPA--- 2278
QY 2518 NDGRPAKRHDIAHSHSPSRLPINSGTWKREHSHKSSSLP-----RVSTWR 2566
Db 2279 ---PNSLDLAPAPAPAPAP-----GLPGDLGDGTLPCEPCTGEL 2318
QY 2567 TGSSSSILSASSSSSKAKSEDEKHVNS---ISGTKQSKENQVS---AKGTWRKIKENEF 2620
Db 2319 TKKPSPLSPSGD--HEANGPGETS LNPPGFVTATAKEEAEAPHAWERGSWPEGAERSS 2376
QY 2621 SPTN--STSGTYSGGATNGAESKTLIYOM-----APAVSKT 2654
Db 2377 RPDULLSSQPLRPGKSSGPPCSSLSEVEAGPQCATDPRPHCGELSPSFNPLPPST 2436

QY 2655 EDVWRIEDCPI-----NNPRSGR-SPTGNTTTPVIDSYSEKANPNIKSDKNQAKQNVGNG 2709
Db 2437 DSDLSLSTEEARLAGGRRRVRGPATGCGPCVAD-----ETPTTSASDSSGSDSDVPPE 2493
QY 2710 SVPMRTVGLNRLTS-----FIQVD-APDQKGT-EIKPGQN-NPVPVSETNESP-----IV 2757
Db 2494 TEECPSTIAEALSDDEGDFLPVDKAGVSGTHHPRGHPPTPLDPRPPRPDVC 2553
QY 2758 ERTFSSSSSS-----KHSSPGTVAARVTPN-----YNPSAKSADSTS- 2799
Db 2554 MADPEGLSSSEGRVERLEKRPORRAPRAKPAAPARRDIRGKSGTTPGKGPVDRISR 2613
QY 2800 --ARPSQIPTPVNNTKK 2815
Db 2614 TVPRSTPSQVTSAEK 2631

RESULT 11
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favellio, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 3.1%; Score 452; DB 2; Length 3507;
Best Local Similarity 19.7%; Pred. No. 1.7e-10;
Matches 458; Conservative 275; Mismatches 859; Indels 738; Gaps 96;
QY 1051 RPXHIIEIKOSEQRQSNQST-----TYPVVTSTDDKHLKFPQHFGQECV 1099
Db 629 RGNHSTSDLIITQVVQQRNFSTGTIILTRGVSSGEAVTQTTDADEF----- 676
QY 1100 SPYRSGANGSETNRVGNHGINQVSQLCEDDDYEDDKPTNYSEYEEQHEE----- 1155
Db 677 -----GLEISAADLAGSGGTLPTT-----LEPKIEGSKKASGGVWTEDEGEDEDIM 726
QY 1156 EERPTNYSIKYN-----EKKRVDP----- 1176
Db 727 EEGSGSWSTINGTGITGSPRSEGTIRVITTLGEDGEPETATKPGISAPDXTGEGSKTE 786
QY 1177 IDYSLKYATDIPSSQKQSFKSSGSGSKTEHMSSESTSTPSSNAKRONQLHPSSA 1236
Db 787 SDGEEKLITVEKDGKEAQSAGSSATSSGKKSEATSGSSSSAKSGTGSEASGSGASSSG 846
QY 1237 QSRSGQPKA-----ATCKVSSINQETIQYCYVEDTPICFRCSLSLSSADEF- 1287
Db 847 SGVSGSGSVSTESGFGTSSGSGVSEATGTGTGDSG-----SGRPSKSTEEKLP 900
QY 1288 -----GQNQTQEDASNT-----LQI 1304
Db 901 FTXNGEKSPISGGDTTGKESSEBETTKPIEGSDSLTEGSGGEWFTGSKGHFEGSKV 960
QY 1305 ABIKGIKGTAEADPVSEVPAVSHPRKTSRRLQSSLSSESARHKAVEFP----- 1355
Db 961 SVTSGKPTQSGAEGSGGPKVPKPGCAPBITTDGESSSTGTGSKGPKADKSNKVN 1020
QY 1356 --SCAKSP-----SKSGAQTPK-----SPPEHYVQETPLMFRCTSVSLDSF 1396

Db 1021 PXTGDKNPDIITDGHEDSTSETSGGQGPKGSKGQPGDGKGEVKKPTSEVDGPQNLSGT 1080
QY 1397 ESRSTASSVQSEPCSGMVGSIISPSDLDP-----SPG-----OTMPPGSR 1436
Db 1081 KGKS-----NVP-----LKPIDLEEGGIILTTSGGKNSTFEHGTKLERLFP----- 1123
QY 1437 KTPPPPPQTAQTKREYRKNKAPTAE-----KREGPKQAAVNAAVQVRVLPDADTILHFA 1493
Db 1124 KTEDKSETPQLGLEISACKPEPEDGTSKEVG-----LEILWESTTPGSTTL 1171
QY 1494 ESTPDGFCSSLSLSLSDLEPIQ-----KDYELRIMPVPVQNDNGNE-- 1536
Db 1172 DSDVGLAISGSDLTKATKKPHVEIGSGTGDBEITATTRDVSKSTKKRVEVDGDNGCE 1231
QY 1537 -----TESQPKESNENQKEAEKTIIDSEK-----LLDDSDDDDIIFILECIIS 1581
Db 1232 TSGVDCKPTTAPTSSSAESSTRIPTTSEASPEGSGEAGVPESPDSGESSTAPDG 1291
QY 1582 AMPTKSRKGGKPAQTAASKLP-----PVARKSQLPVYKLLPSONELQPKHVSFTP 1634
Db 1292 VSPTSATAPVETTSASSTPDAVESSGIPSTSKPTAREPLETTAPTEVTSPEG--SGTE 1349
QY 1635 GDDMPRVYCEGT-PINFSTATSLDITTESPNELAAAGEVR-----GGAQSGE 1683
Db 1350 ESTLPP-----TEGSGESTTSAPTVEPATVLPQNRNEKEPPTKDTFALPTTTTGAPOAND 1406
QY 1684 FEKRDITPTEGRSTDEA-----QGG-KTSSVTIPE----- 1712
Db 1407 SSVENTKCT---SSDEGLDLCERRTGVRCRCEPFGGAPPKKSCVDVDECATGDHNCHE 1463
QY 1713 -----LDNKAEB-----GDLAEICIN 1729
Db 1464 SARCONVGVGACFPTGFRKADDDGCGQDIDECTEHNSTCCGANAKVKNKPGTVSCECEN 1523
QY 1730 SAMPKKGSHKPRFRVKIMDQVQASASSAPNKN-QLD-----GKXKKTSPV 1776
Db 1524 GFLGDGVQCVF-TTKKPCDSTQSKSHCESNMSCEVDTVDGSGVECKEKGKYGKVKVC 1582
QY 1777 KPIONTEYTRVRKNADSKNNLNAERVSDNKKQNLKNS--KDFND-----KL 1827
Db 1583 EDINECVAEKAPCSLNA---NCVNMNGTFS---CSCKQGRGDFGFMCTDINECDERHPCH 1636
QY 1828 PNNE-DRVRSGFAFDSPHY-----TPIBGT-----PYCFS--RND 1861
Db 1637 PHAECTNLESGFCECHSGFEGDGIKKTNPLERSCEDVEKFCGRVDHVSCLSVIYNGS. 1696
QY 1862 LSSL-----DFDDDDVLSREKAEKLRKA-----KENKESEAK 1893
Db 1697 LSSVCECEPGFRFEKESNSCVDIDECESRNNCDPASAVCVNTEGSRVCEAGYEGEGG 1756
QY 1894 VTSHELTSNQOSANKTOAIKQPINR-----GQPKPILOKQSTFPQSSKDI 1940
Db 1757 VCTDIDECDRGMAGCDSNMC---INRYSGCGCKMAGYTGDCGATCIKIEE-PKS---- 1808
QY 1941 PDGAATDE-----KLQNFALNTPVCFSS-----H--NSLSL--SDIDQENN-- 1980
Db 1809 -DKTACTDENSRLCELEKKQCTVDEEVPOCGACLPGHHPINGTCQSLQISGLCAQKND 1867
QY 1981 NKEPEPIKETEPPDSQGEFSPKQASGYAPKSFHVEDTPVCFSRNSSLSSLSIDSEDLQ 2040
Db 1868 NKHAECI--DIHPDSH-----FCSCPQGF-IGDGMICDDVDECNAGMCDDEKCE 1916
QY 2041 ECIS-------AMPKKKPSR--LKGDNEKHSPPNMG-----GILGE- 2074
Db 1917 NTIGSFNVCLEGFKKVDEKCVDEKKQPNREKIEIDENSSSSNSGQKPTTKGIVSST 1976
QY 2075 DLTLDKDIQPDSEHGLSPSENFDMKATQEGANSIVSSILHQAAAACLSRQASSSDS 2134
Db 1977 SATSESTTAEPHVTTTSSITSTTKDMTSSKSPENVMTSS--ESPFEVTSKSTTASSET 2034
QY 2135 ILSLKSGLSLGSPFHLTPDQ--BEKPTSNKGR--ILKPGKSTLETKKIESKGI 2188
Db 2035 TVS-----STFESSSSSEAPLTSPPATTEVITSSSVKST--TPKESSSE-- 2078

941	Db	ASTSASEGASTSASAGASTSASAGASTSASAGASTSASAGASTSASAGAST--SASASAS	998
1448	Qy	TREVPKNKAPTAEKREGPKQAANVAOVORVQLPDADTLLHPATESTPDGFSCSSLS	1507
999	Db	TSASAGASTSASAGASTSASAGASISAS-----ESASTSASAGASTSVASATSAS	1050
1508	Qy	ALSLDPEFFIOKVLEIRMPVQENDNGNETESOPKESNENQBEAKIIDSEKDILDDS	1567
1051	Db	A-----SASTSAGESASTSASAGASTSASEASTSASASASTSAS	1090
1568	Qy	DDDDIBTEBECIISAMPTKSRKGKPAQTASKLPPPVAREKPSQLPYVKLLPQNRLQP	1627
1091	Db	ASASISAGESASTSASAGASTSASAGASTSASASASTSA-----SESASTST	1137
1628	Qy	KHVSFTPGDMPRVYCVGEPTPIFNSTATSLDTITESPNNELAAEGVRGAOSGBEKR	1687
1138	Db	SAGASTSASES-----ASTSASAGASTSAS-----ASASTSASAGASTSASAGASTSASA	1187
1688	Qy	DTIPTGRSDTAQGCKTSSVTIPELDNDKAEBGDILAECINAMPKGKSHKPFVKIM	1747
1188	Db	STASGESTSASAGASTSASAGASTSASAGASTSASAGASTSASASTS-----	1238
1748	Qy	DVOQOASASSAPNKQLDGKKKPTSPVKPIPONTEYRTRVRKNADSKNLNARVFSD	1807
1239	Db	--ASASASTSASAGASTSASAGASTSASESASTSASAGASTSASAGASTSASAGASTSA	1295
1808	Qy	NKDSKQNLKNNGKDNFLPNNEDRVGRGFAFDPSPHHYPIEGTPVCFSERNDSLGLDF	1867
1296	Db	SAGASTSVASASTSASESASTSAS-ASASTSASAGASTSAGESTSASAGASTSGASES	1354
1868	Qy	DDDDVLRSKEBALRKAKENKSEAKVTSHTELTSNOQSANKTQAIKQPINRGQPKPIL	1927
1355	Db	ASTSAGASTSASAGASTSASAGASTSASAGASTSASAGASTSASAGA-----	1401
1928	Qy	OKOSTPPOSKDIHDRCAATDEKLQNPAINENTPVCFSHNSLSLSLSDIDENNKNENEPI	1987
1402	Db	SASTSASAGASTSASAGASTSASAGASTSASAGASTSASAGASTSASAGASTSASASASI	1461
1988	Qy	KETEPDPSOGEPS-KPOASGYAPKSFHVE-DTPVCFSRNSSL-SLSDIEDLLLOECIS	2044
1462	Db	SASESASTSASAGASTSASAGASTSVASASTSASAGASTSASAGASTSASESASTSASA	1521
2045	Qy	SAMPKKKPERLKGDNEKHSPRNWGGJLGDBDLTLDLKDIORPOSEHGLSPDSNFOWKAI	2104
1522	Db	SASTSASAGASTSASAGASTSASAGASTSASESASTSASAGASTSASAGASTSASAGAST	1581
2105	Qy	QEGANSTVSLHQAAAACI-SRQASSDSDSIILSKGIGSLGPSPHHTPDQEEKPFTSNKG	2164
1582	Db	SAGASTSASAGASTSASAGASTSASAGASTSAGASTSASAGASTSASESASTSASA	1641
2165	Qy	PRILKPGEKSTLETTKIESKGIKGKKVYKSLIT---OKVRSNSEISQMQRQLOANNM	2221
1642	Db	SAGTASAGASTSAGASTSASVSASTSASESASTSASAGASTSASAGASTSASAGAST	1701
2222	Qy	PSISRGETMIHIQVNRNSSASTSPVKKGPLKTPASKPSSEBOTATTSPRGAKPSVKSE	2281
1702	Db	SAGEASTSAGASTSASAGASTSASESASTSAGASTSASAGASTSASAGASTSASA	1761
2282	Qy	LSPVAROTSIQTGSSKAPSRSGRDSTPSPRAQOLSRPIQSPORNISIPCRNGISPPNKK	2341
1762	Db	SAGASTSAGASTSASAGASTSAGASTSASAGASTSASESASTSASESASTSTS-----	1816
2342	Qy	LSQLPTRSPSTASTKSGSGKMGTYPGRMQSQONLTQOTGLSKMNASTIPRESASKGL	2401
1817	Db	ASESASTSAGASTSASAGASTSASA---SASTSASTSASESASTSAGASTSASTSAS	1872
2402	Qy	NOMNINGANKVBLSMSGTKSGESDESERFVLVRQGTTFIKEAESPTRLRKLBEBSA	2461
1873	Db	ASASTSAGASTSASAGASTSASAGASTSAGASTSVASASTSASAGASTSASESAS	1932
2462	Qy	FESLUSPSRRPASPTRSOAQTPVLSPSLPDMGLSTHSVQAGWRKLPNNLSPTIEYNDGR	2521
1933	Db	TSASAGASTSAGASTSASAGASTS-ASASASTSAGASA-----SASTSAGASTS	1971

Qy	2522	PAKRHDIARSHSTSPGRPLPINRGTWKREKSHKSSSLPRVSTWRRRTGSSSSIIISASSESS	2581
Db	1972	-----STGASESTASASASTSASASASTSASASASTSASASASTSASASASTSASASASTS	2024
Qy	2582	EKAKEDEKVNISIGTKQSKENQVAKGQWRIKENERPPTNSTOTVSSGATNGAESK	2641
Db	2025	ESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTS	2084
Qy	2642	TLIYQMAPAVSKTEDVWVRIEDCPINPRSGRSGPTGNTPPVIDSVSEKAMPNIKDSKDNQ	2701
Db	2085	T-----SASASASTS-----ASASASTSASASASTSASESASTSASASASTS	2126
Qy	2702	AKQNVGNGSVPMRTVGLNRLTSFIQVADPDQXGTEIKPQNNPVPVSETNESPIVERTP	2761
Db	2127	ASESA-----STSASASTSASASTSASASASTSASASASTSASASASTS	2166
Qy	2762	FSSSSGSKHSPGCTVAARVTPFNNPSPKKSADSTASRPSQIPTPVNNNTKKRDKGKTD	2821
Db	2167	ASASASTSASASTSASASTSASASTSASEGASTSASASTSASASASTSASASA-----	2220
Qy	2822	STSSGTSQSPKRHSGS	2837
Db	2221	ASASASTSASASASTS	2236
RESULT 13			
T20532			
hypothetical protein F07A11.6b - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999			
C:Accession: J20532; J27777			
R:Palmer, S.			
submitted to the EMBL Data Library, October 1995			
A:Reference number: J19287			
A:Accession: J20532			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-2722 <W1>			
A:Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b			
A:Experimental source: clone F07A11			
R:Gajadaty, S.			
submitted to the EMBL Data Library, March 1996			
A:Reference number: J20417			
A:Accession: J27777			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-2722 <W12>			
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b			
A:Experimental source: clone ZK20			
C:Genetics:			
A:Gene: CESP:F07A11.6b			
A:Map position: 2			
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3;			
Query Match 2.9%; Score 428.5; DB 2; Length 2722;			
Best Local Similarity 18.6%; Pred.No.1.le-09;			
Matches 460; Conservative 324; Mismatches 893; Indels 801; Gaps 107;			
Qy	811	SDNFGNTMTVLSPLYNTTVLPSSSSRGSLDSRSRKSRLERIGIGLGNHYHPATENP	870
Db	357	SDRNL-----YFINRKNENTEVARSSPRTSKSENDQ-----GSSSP-----	394
Qy	871	GTSSKRGLOISTTAAQIAKVMVEVSAIHTSQEDR---SSGSTTELHCVTDERNALRRSSA	927
Db	395	--SSSRDRQLNLDPLQTRSSVEH---HTNQDOENNASGDS-----SSDSDEEGSSS	443
Qy	928	AHTSNNTYNTKSNNSRNTCSMPYAKLEYKRSNDLSNVSSNDGYGKRGQMKPSIESYS	987
Db	444	SNESDSQNDVDEDDDDVVS-----FEKR--HEPEGKSSSPGNGHRDESNG--DKDH	493
Qy	988	EDDESKFCSYGQYPADLAHKIHSANHMDNDGEL-----DTPINYSLKYSDEQLNSGRQS	1042

QY 2758 ERTFSSSSSKHSSPS---GTVAARVTFNVPNSPRKSSADTSARPSQIPTPVNNT 2813
Db 2428 KOIPTSEEDTDDSKADSMCAEAGAFRIL-----SRSTMGNSGSPASGT----- 2474
QY 2814 KRDSTDTSTESGTSQSP 2831
Db 2475 --TSPSTSSSISGSPDSP 2490

RESULT 14
E88320
Protein F07A11.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: AF5000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/usc/c_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2738 <STO>
A:Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A11
C:Genetics:
A:Gene: F07A11.6
A:Map position: 2

Query Match
Best Local Similarity 18.5%; Pred. No. 1.4e-09;
Matches 461; Conservative 325; Mismatches 891; Indels 817; Gaps 108;

QY 811 SNFNTGNMTVLSFYINTVLPSSSSRGSLDSRSKXDRSLERGIQGLNYHPATENP 870
Db 357 SDRFNL-----YFINKKENIEVAARSSPTSSENDQ-----GSSSP----- 394

QY 871 GTSSRGQIQTAAQIAQIAKMEVSAIHTSQDR---SSGSTTEHCVTDNRALRRSA 927
Db 395 --SSSRDRQNLHDLPTQTRSSVEH-----HTNQDQENNASGDS-----SSDSDEEGSS 443

QY 928 AHTHTNTVFTKSNRNTSCMPYAKLEYKRSNDSLSNSNDGCGYKRGOMKPSIESYS 987
Db 444 SNEDSDQNDVDEEDEDVVS-----BEKR--HEPEGKSSSPGNGHDSNG--DKDH 493

QY 988 EDESKFCSGYGPADLAHIAHNSAHNDNDGEL-----DTPINYSKYSDQLNSGRQS 1042
Db 494 EDSSERF---SQPSTSSH--HETSHSPKDSSEAYQSRSFPLNY-----QS 534

QY 1043 PSQNERWARPKHIIEDETKQS--EQCRSNQSTTYPVYTESTDD-----KHLKQPHEGQ 1095
Db 535 QSPGYEFLEK-----EIQEFSPTSSASSDLEDMFPDNLTRMLERNHWRPFID- 588

QY 1096 QCVSPYRSRGANGSETNRVGSNHGINQVQSGL-----CQEDDYED-----DK 1139
Db 589 ---VSSF-----VNRIDBEIVELNKAASVEKFTGRPPKCNNDVLSIQIIVFHE 636

QY 1140 PTNYERYSEEBQHEERPTNYSKYNEKHVDQPIDYSLKATDIPSSQKQSFPSK 1199
Db 637 PRDY--YYENPCSELEVRIDW-----RKLSDTADLD--DFRATDSKELGRDQ 681

QY 1200 SSSGQSKTEHMSSSSENTSTPSSNAKQNL-----HPSSAQSRSQCPKAATCKVSS 1253
Db 682 PAGGITS-----GRPSLDESRTNLSFPDTHHPALQRS----- 716

QY 1254 INQETIQYCV-----EDTPICFSRCSLSLSAEDIGCQNTQOEDASNTLQI----- 1304
Db 717 -----HSLCIGPMTPTTF-----PTSQPLLVNTTHLPQTSG 748

QY 1305 AEIKGIGTGAEDPVSVPAVSOHPRTKSRLOGSSLSSESARHKAVEFPSSGAK----- 1359
Db 749 PSTSGIITPRSSQPPPLMSPVSRNSMSST---GRPASIQTLRHQSVMFPDPVSIPIPPP 805

QY 1360 -----SPKSGAQT-----PKSPPEHVHVBETPLMFSR 1386
Db 806 IPPTHDEMAMPRTGPPRRSRSETMVLRLSPFPCTPIQNLTMPIVPPPHLIAAT----- 859
QY 1387 CTSVSSILDSFESRSIASSVQSEPCSGMVSGIISPSDLDPSPGOTMPPSRKTPPPPPQTA 1446
Db 860 -----STGTHSVSSSAHSTP-RHSISG-----TPVH-CEFSNKTSCQP--TP 898

QY 1447 QTKREVPNKAPIAKRESGPKQAAVNAVQVQLPDAITLLHFATSTPDGFSCESSL 1506
Db 899 KSRPEKQIIRHDTISK--SGFSN-AINALQARSQMSGDP--KKSAPSTPVVRDAGSGL 953
QY 1507 SALSLEDPFTQKQVELRIMPPVOE-----NNGNET----- 1537
Db 954 VAQIMSN---QPNLGLRLKLPRIEKKSSALQNIQHQPPhNSANSTPSTPSTHQAOMPKD 1010

QY 1538 ESQPKESNENOPKEAB----- 1554
Db 1011 KEKERKKEKEKERERAREREMKRKETKEERNKEMERAKLEDERQERKEKKERD 1070
QY 1555 -----KTIDSEK-----DLDDSDDDIEILEECIIISAMPTKSSRKGGKPA 1595
Db 1071 ERKEKEKVRKKAKEKELKKKKHKKGSDSDSDNDELDDLVRKS--TKEMTQEEKDH 1128

QY 1596 QTASKLPPPVARPKSPQLPVYKLLPSONRLOPKXHVSTPGDDMPRVYCVVEGTPINFSTAT 1655
Db 1129 QLALL-----SKGIILNKSRERSDKRAHDSFE-----KMQQKS 1164

QY 1556 SLDLTIESPPNELAAGEVIRGAOSGE---FKRD-----TTPTEGRSTDE---AQGG 1703
Db 1165 QORVLIESDDE--GGKDGDKGNSNGEESDSEKADLPPPPAPFSLSESADQRLKVLKER 1223

QY 1704 KTSVITPELDNKAKEGDILAEICINSAMPKSHKPPRVKIMDOVQOASASSAPKN 1763
Db 1224 EKGLTITSSDEHDHAGELHQQLTREDNRKQKSLTAYSSDEQGERKNVFRWRDD 1283

QY 1764 QLGKKKKKTPSVKPIQNTTEYTRVRKNADSKNNLNAERVSFSD-----NKDSK 1812
Db 1284 SEDAAAKHPGWSAKDQKQRKLEHRRSEDESKNAKDRFDIPIHEDVSDSEETEDGS 1343

QY 1813 KQNLKNSKDFNDKLPNNEDVRVG-----SFAPDSPHYHTPIEGPYCF 1856
Db 1344 RSRQSTSTISNTVAKERKESGKPLRIVPEPTGTPLSPKILSPKXLSPTST--SS 1401

QY 1857 SRNDSJSLDFDDDDVLSREKAEALRKAKEKSEAKVTSHLTSLTNQSQSANKTOAIKQ 1916
Db 1402 TKRSSIS---DHENLISPRQ-----NRTTSSTST-TATTSSKHEALSIEPK 1443

QY 1917 PINRGQPKPILOKSTFPQSSKOIP-----DRGAATDEKLQNF-----AIEN 1958
Db 1444 PLS-----PPVTAKSSV---SSIDDPISIRDEFSSMNSAADSMPMTTGRPMVLTKAMKAFNS 1496

QY 1959 TPV-----CF-----SHNSSLSSLSIDID-----QENNKENEPIKET 1990
Db 1497 TPKKYVSYSLLIDCYMLGMWMAKNSSGQHDSSGSSSSSSSDSGSTSSDSDSDEVPKQT 1556

QY 1991 EPDPS-----QGEPSK-----PQASGVAPKSFHVEDTPVCFSRNSSLSSLSIDREDDL 2038
Db 1557 EPVTSIPVWASDNGSPENVVETFSVSQTPR---EPFPFTISEQS-----ESEPEA 1606

QY 2039 LQSC-ISSAMPKKKPSRLKGDNEKHSPRNMGILGEDTLDLKDIQRPDSEHGLSPDSE 2097
Db 1607 VPQCPASVEPQNETSQNTVEPVSEEHDSHEHG-----DSEAVESQOQP-LEH--QEEKE 1659

QY 2098 NFDWKAIQEGANSIVSSLHQAAAAACLRSQSSDSLSLSKSGISLGS--PFHLTPDQE 2155
Db 1660 ELENKILD-----VAAHEHEEQVQGEDSDV--BSSIPAPSDPEDPVTOAQE 1703

QY 2156 EKFPET-----SNKGPRLLKQCE----- 2172
Db 1704 KSAHTLISDQETQAVQSIFDEEEADEFPQYDFGISTNEKEVSGKDPHNKITEPLANG 1763

QY 2173 -----KSTLETKKIESBKIGKGGKV-YKSLITCKVRNSSEISGQMK 2214
 DB 1764 HTDLLSPSSSAHASEKQSTKSEDDHEEDSELVMEKEVPMQVIAQEVHVPSESPMEE 1823
 QY 2215 Q-----PQANWPSISRGRTMTHIPQVRNSSSSTSPVSKGKPLKT 2255
 DB 1824 EVKLETSFVPEKBEFKMEESFEQPTFDLISNNESQDTPGAVNNH-----LHENHDAVOT 1878
 QY 2256 PASKSP-SEGQTATTSRGAQKPSVKSELSVAVQTSQIGGSSKAPSRSGSRDSTSPRPAQ 2314
 DB 1879 PIQLOPASQVQAPSRPA-----VAPDSQONGPV-----LVSQSQSPS---- 1918
 QY 2315 QPLSRPTQSPGRNISPRNGISPPNKSQLPRTSSPSTASTKSGSGKQVYTSRQMS 2374
 DB 1919 -PMSQOSQDMAQNLLSSKDINDLAALKHKNPEALA---QATRGDCSGIFQHLHHAQGN 1974
 QY 2375 QONLTKQTGLSKWASSTIPRESASKGLNOMN-----NGNGANKKVELSRMSSTKSSGSE 2428
 DB 1975 GQNTPEMLQKAAFAAQOQENANQWQAKMKQQTINKDRIKEQERVYKYEENKVE 2034
 QY 2429 SDRSER-----PVLVRQSTFIKEAPSPTRLRKE-----ESAS-- 2461
 DB 2035 EDREKQKBEERQELAAATAATATQKAAEALKQKQEVPRHGFQHLVSMMTPEARSILY 2094
 QY 2462 --FESLS-----PS---SRPASPTRSQAQTPVLSPLDPKSLSTHS-- 2497
 DB 2095 EQFPGLSSYINRDSIGATNGVLHLPTQSIQRPSS--TASTSSNPPKAPLOPASVNONITD 2153
 QY 2498 -----SVQAGWRKLPNLSPTIEYNDGRPAKHDIARSHSESPSRLPINRSQTKWR--- 2549
 DB 2154 PAETEEIRVQRFYKPLKMSAE-----EAATMAVASSDPNPATSTVDLAAMLQOLA 2207
 QY 2550 -EHSKHSLSIPRVSTWRTGSSSI--LSASSSESEKAKSEDEKHVNSIGTKQSKENQV 2606
 DB 2208 AQAQAQAQVPPVVTASTPNPLNLETLLTASLANLATGALNPLSLMALTSSLNQOSP 2267
 QY 2607 SAKGTWRKIKENESPTNSTSTQVSSGATNGAGSKTLIYOMAPAVSKTEDVWVRIEDCPI 2666
 DB 2268 VYQIARVLLTMWQMLATHQISELLATWN-QEETLMALLAR-----NGLFF 2315
 QY 2667 NNPRSGSP---TCN--TPPVIDSVSEKANPNIKOSKONQAKONVONGVPMRTVGLN 2720
 DB 2316 AMPQONQOPQMPAQGGAFAITVPLPHMSLKEN-----AKD---QLSVGGVSDRKSCPLHA 2367
 QY 2721 RLTSFIQVDAPDQGTIKG-----QNNPVPVSETNE----- 2753
 DB 2368 MIGGQOQPPPPQPMQAVAPAPRSPSPRKSFMFENLPPMEKKNEMFRKEILRLDIL 2427
 QY 2754 -----SPIVERTPFSSSSSKHSPPS-----GTVAARVTPPNVNPSPRKSADS 2797
 DB 2428 LEEELGADEEDQKDLAQIPTSEEDTODSKADSMGABGSAFRIL-----SRSTMGN 2480
 QY 2798 TSARPSQIPTPVNNTKRSKTDSTSSSGTQSP 2831
 DB 2481 NSGSPSASGT-----TSPTSSSSISSGPDSP 2506

RESULT 15
 T02345
 Hypothetical protein KIAA0324 - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C:Accession: T02345
 R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 submitted to the EMBL Data Library, March 1998
 A:Description: Sequencing of human chromosome 16p13.3.
 A:Reference number: Z14664
 A:Accession: T02345
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1791 <RIC>
 A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650

C:Genetics:
 A:Map position: 16
 A:Introns: 1610/2; 1706/2
 A>Note: KIAA0324
 Query Match 2.9%; Score 419.5; DB 2; Length 1791;
 Best Local Similarity 19.6%; Pred. No. 1.5e-09;
 Matches 374; Conservative 259; Mismatches 742; Indels 537; Gaps 75;
 QY 1210 HSSSSSENTS-----TPSSNAKRNQLHPSSAQSRSQOPQAATCKVSSI 1254
 DB 1 HSGSSSDTKVKPETPRQSHSGSISPVKVAQTTPGSLSGSKSPC----- 49
 QY 1255 NOETIQYCVEDTIFICRSRSSLSSAEDICGNTTQEAADSANTLQIAETLKGIGTR 1314
 DB 50 -QEKSKDSLQVSCFGSLCA-----GVKSSTPPGESYFVGSLLQKGO--SQ 94
 QY 1315 SAEDPVEVPVAVSQHPRTKSRLOGSSLSSESARHKAVEFPFGAKSPSKGCAQTPKSPPE 1374
 DB 95 TSPHRSD-----TSPVQRQSHSESPSLQSKS-----QTSPPKGRSRSSSPVT 138
 QY 1375 HYVQETPLM-----FSRCTSVSSLDSPESRSIASVQSEPCSGMVSGIISPSDLDPSPGQ 1429
 DB 139 ELASRSPTRQDRGFSAAPMLKSGMSPEQSRFQSDSSSYFTVDS-NSLLQGSRLTAEAK 197
 QY 1430 ---TWPPSRSKTPPPBPQT-----AQTKEVPKNKAPTAEKRESGPK 1468
 DB 198 EKMLAPPQEDATAPPKOKFSPFPVQDRPESSLVFKDLRTPPRERSGAGSPETKEQ 257
 QY 1469 QAAVNAARVQVVL-----PDADTLHFAT--ESTPDGFCSSSL-----ALSLD 1512
 DB 258 NSALPTSSQDEELMEVVEKSEEPAGQILSHLSSELKEMSTSNFESSPEVEERPAVSLTD 317
 QY 1513 EPFTQKQVLEIMP-----PVQENDNGN--ETESEQPKESNENQ 1549
 DB 318 QSQOASLEAVEVPSMASSWGCGPHFSPEHKELNSPLRENSFGSPLFRNSGGPLCTEWT 377
 QY 1550 EKEAKTTDSEKDLDDDDDDIIELECIISAMPTKSSRKKKPAQTAGSKLPPPVARKP 1609
 DB 378 GFSSEVKEDLNGPFLNQLTDPSLDMKE---QSTRSS-----GHSSELSPDAVEKA 426
 QY 1610 SOLPVYKLLPQNRLQPOKHVSFTPGDMPRVYCVGPTPNFSTATS--LSDLTIESPPN 1667
 DB 427 G-----MSNQSISSPVLDAVPTPRERS-----SSASSPEMKDGLPRTPSR 469
 QY 1668 ELAAGE--GVGGAGSGFEKRDPTTTEGRSTDEAQGGKTSVVTIPBLDNKABEGDILA 1725
 DB 470 RSRSGSPGLRDGS-----GTPSRHSLSGSS-----PGMKDIPRTPSRGRS 510
 QY 1726 ECINSAMPKGSKHKPFRVKTMQVQASASSAPNKNQ---LOGKKKKPTSPVKPIQNT 1783
 DB 511 ECDSSPEPKALPQT-----RPRSRSPFELNNKCLTPQRRSGSE--SSVDQKT 559
 QY 1784 EYRTEV--RKNAKSNLNAERVFSDNKKONLKNNSKDFNDKLPNNEDRVRGSPAFD 1841
 DB 560 VARTLQGRSRSGSQELDVPSASQPSRSE-----SDSSPDSKAKTR----- 602
 QY 1842 SPHYHTPIEGTYCFYCFRNDLSLSDFFDDDDVDLGRKAELEKAKENKESAKVTSHTLT 1901
 DB 603 -----TPLR-----QSRSGSGSPEVD-----SKRSLSPRRSRSGSPSEVDKPRAA 643
 QY 1902 SNOQKANTQAIKAPINRGOPKPILOKQSTFPQSSK---DIPRGAATDEKLNQFAEN 1958
 DB 644 PRAQSGSSSEPKAPAPRALP-----RRSRSGSSKGRGSPGSGSSTES-----SPEH 693
 QY 1959 TPVCFSHNSLSLSDIDQENNNKENEP-IKETBPDSQGPSPQASGYAPKSFHVED- 2016
 DB 694 P-----KSRTARRGSRSPPEPKTSRTPPRRSRSSRSPFELTRKARLSRRSRA 742
 QY 2017 -----TPVCFSRNSLSLSDSDDLLQBCISSAMPKPKKPKRLKDNKESPRN 2067
 DB 743 SSSPETSRTPPHRRSRSPSVSPFAEKSRSSRRRRSSSPRTKTTTSR-RGRSPSPKPRG 801

Search completed: August 25, 2004, 17:24:44
Job time : 98 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:21 ; Search time 32.5 Seconds

(without alignments)
4554.937 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575
Sequence: 1 MAASYPQLLKQVEALKVEN.....ESSGTQSPKRHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14533	99.7	2843	1 APC_HUMAN	P25054 homo sapien
2	13168.5	90.3	2842	1 APC_RAT	P70478 rattus norv
3	13148	90.2	2845	1 APC_MOUSE	Q61315 mus musculu
4	477	3.3	2774	1 MAPA_RAT	P34926 rattus norv
5	474.5	3.3	5560	1 SPEN_DROME	Q88x83 drosophila
6	445	3.1	3664	1 MINT_HUMAN	Q86t58 homo sapien
7	434.5	3.0	3644	1 MINT_MOUSE	Q62504 mus musculu
8	417	2.9	2688	1 ZEP1_MOUSE	Q03172 mus musculu
9	400.5	2.7	2805	1 MAPA_HUMAN	P78559 homo sapien
10	400	2.7	5085	1 PCLO_RAT	Q9Jks6 rattus norv
11	395	2.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
12	387.5	2.7	5147	1 PCLO_HUMAN	Q9Y6V0 homo sapien
13	385.5	2.6	3210	1 CENF_HUMAN	P49454 homo sapien
14	381	2.6	2459	1 MAPB_RAT	P15205 rattus norv
15	380.5	2.6	4377	1 ANK3_HUMAN	Q12955 homo sapien
16	379.5	2.6	4911	1 MLL3_HUMAN	Q8ne24 homo sapien
17	375.5	2.6	3562	1 PCLO_MOUSE	Q90953 gallus gall
18	374	2.6	5038	1 PCLO_MOUSE	Q9QYX7 mus musculu
19	373	2.6	3256	1 K167_HUMAN	P46013 homo sapien
20	371.5	2.5	3969	1 HRX_HUMAN	Q03164 homo sapien
21	371	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
22	369.5	2.5	2464	1 MAPB_MOUSE	P14873 mus musculu
23	368.5	2.5	2453	1 NCR1_MOUSE	Q60974 mus musculu
24	368	2.5	2440	1 NCR1_HUMAN	Q75376 homo sapien
25	366	2.5	2492	1 ATRX_FANTR	Q7Yqm4 pan troglod
26	365	2.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
27	365	2.5	3866	1 HRX_MOUSE	P55200 mus musculu
28	365	2.5	5120	1 PCLO_CHICK	Q9pu36 gallus gall
29	364.5	2.5	1581	1 PRB_HUMAN	Q15648 h peroxisom
30	362	2.5	3259	1 GBL1_HUMAN	Q14789 homo sapien
31	360.5	2.5	2492	1 UNR9_CAEEL	Q01761 caenorhabdi
32	357	2.4	2492	1 ATRX_PONPY	Q7Yqm3 pongo pygma
33	356.5	2.4	2663	1 CENE_HUMAN	Q02224 homo sapien

ALIGNMENTS

RESULT 1

ID	APC_HUMAN	STANDARD;	PRT;	2843 AA.
AC	P25054; Q15162; Q15163;			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Adenomatous polyposis coli protein (APC protein).			
GN	APC OR DP2.5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91335210; PubMed=1651562;			
RA	Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,			
RA	Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,			
RA	Finleary R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,			
RA	Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,			
RT	"Identification of FAP locus genes from chromosome 5q21.";			
RL	Science 253:661-665(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=91330307; PubMed=1678319;			
RA	Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,			
RA	Samowitz L., Groden J., Stevens J., Spirio L., Robertson M.,			
RA	Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,			
RA	Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,			
RA	Abderahim H., Cohen D., Leppert M., White R.,			
RT	"Identification of deletion mutations and three new genes at the			
RT	familial polyposis locus.";			
RL	Cell 66:601-613(1991).			
RN	[3]			
RP	ASSOCIATION WITH CATENINS.			
RX	MEDLINE=94082295; PubMed=8259519;			
RA	Su L.-K., Vogelstein B., Kinzler K.W.;			
RT	"Association of the APC tumor suppressor protein with catenins.";			
RL	Science 262:11734-11737(1993).			
RN	[4]			
RP	DISEASE.			
RX	MEDLINE=95174843; PubMed=7661930;			
RA	Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,			
RA	Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,			
RA	Wood P.A., Taqi F., Bocker S.V., Petersen G.M., Offerhaus G.J.A.,			
RA	Tersmette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;			
RT	"The molecular basis of Turcot's syndrome.";			
RL	New Engl. J. Med. 332:839-847(1995).			
RN	[5]			
RP	DISEASE.			
RX	MEDLINE=97094176; PubMed=8940264;			
RA	Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,			
RA	Fisher A., Barber J., du Boulay C., Primrose J., Burn J., Fodde R.;			
RT	"Hereditary desmoid disease due to a frameshift mutation at codon 1924			
RT	of the APC gene.";			

Q62059 mus musculu
P46100 homo sapien
Q912u6 mus musculu
P08640 saccharomyc
P30415 mus musculu
P30414 homo sapien
Q12888 homo sapien
Q9wu42 mus musculu
P47035 saccharomyc
Q9nm4 caenorhabdi
O35923 rattus norv
P48415 saccharomyc

RL Am. J. Hum. Genet. 59:1193-1201(1996).
 RN [6]
 RP DISEASE.
 RX MEDLINE=20243021; PubMed=10782927;
 RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
 RA Fodde R., Alman B., Bapat B.;
 RA "A germline mutation at the extreme 3-prime end of the APC gene
 RT results in a severe desmoid phenotype and is associated with
 RT overexpression of beta-catenin in the desmoid tumor.";
 RL Clin. Genet. 57:205-212(2000).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
 RX MEDLINE=20384842; PubMed=10926498;
 RA Day C.L., Alber T.;
 RA "Crystal structure of the amino-terminal coiled-coil domain of the
 RT APC tumor suppressor.";
 RL J. Mol. Biol. 301:147-156(2000).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
 CTNNB1.
 RX MEDLINE=21564054; PubMed=11707392;
 RA Eklof Spink K., Fridman S.G., Weis W.I.;
 RA "Molecular mechanisms of beta-catenin recognition by adenomatous
 RT polyposis coli revealed by the structure of an APC-beta-catenin
 RT complex.";
 RL EMBO J. 20:6203-6212(2001).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
 AXIN.
 RX MEDLINE=20271867; PubMed=10811618;
 RA Spink K.E., Polakis P., Weis W.I.;
 RA "Structural basis of the axin-adenomatous polyposis coli
 RT interaction.";
 RL EMBO J. 19:2270-2279(2000).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94154728; PubMed=8111410;
 RA Nagase H., Nakamura Y.;
 RA "Mutations of the APC (adenomatous polyposis coli) gene.";
 RL Hum. Mutat. 2:425-434(1993).
 RN [11]
 RP VARIANTS FAP.
 RX MEDLINE=91355211; PubMed=1651563;
 RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
 RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
 RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
 RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
 RA "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.";
 RL Science 253:665-669(1991).
 RN [12]
 RP VARIANTS FAP.
 RX MEDLINE=93265030; PubMed=1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.;
 RA "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.";
 RL Hum. Mol. Genet. 1:229-233(1992).
 RN [13]
 RP VARIANTS FAP.
 RX MEDLINE=93244793; PubMed=1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.;
 RA "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.";
 RL Hum. Mol. Genet. 1:559-563(1992).
 RN [14]
 RP VARIANT FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
 RP SER-2502
 RX MEDLINE=93250848; PubMed=1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,

RA Baba S., Nakamura Y.;
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RT patients: 61 new patients and a summary of 150 unrelated patients.";
 RL Hum. Mutat. 1:467-473(1992).
 RN [15]
 RP VARIANT FAP TRP-99.
 RX TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=95134544; PubMed=7833149;
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.;
 RA "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.";
 RL Eur. J. Cancer 30A:1709-1713(1994).
 RN [16]
 RP VARIANT FAP GLY-722.
 RX MEDLINE=9513430; PubMed=7833931;
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pillia S., Prete F., Mareni C., Guanti G.;
 RA "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.";
 RL Hum. Mol. Genet. 3:1687-1688(1994).
 RN [17]
 RP ERRATUM.
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pillia S., Prete F., Mareni C., Guanti G.;
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [18]
 RP VARIANT FAP ILE-171.
 RX MEDLINE=97144176; PubMed=8990002;
 RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Correlisse I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.;
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
 RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGGE, PTT, and southern analysis.";
 RL Hum. Mutat. 9:7-16(1997).
 RN [19]
 RP VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
 RX MEDLINE=98080146; PubMed=9419979;
 RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
 RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
 RA Fukutome A., Tomiyama J., Chuganji Y., Momoi M., Utsunomiya J.;
 RT "Drastic genetic instability of tumors and normal tissues in Turcot
 RT syndrome.";
 RL Oncogene 15:2877-2881(1997).
 RN [20]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400248; PubMed=9731522;
 RA Redston M., Nathanson K.L., Yuan Z.-Q., Neuhausen S.L., Satagopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
 RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
 RT "The APC I1307K allele and breast cancer risk.";
 RL Nat. Genet. 20:13-14(1998).
 RN [21]
 RP VARIANTS LYS-1307 AND GLN-1317.
 RX TISSUE=Peripheral blood;
 RX MEDLINE=98393712; PubMed=9724771;
 RA Frayling I.M., Beck N.E., Ilyas M., Hodgson S.V., Thomas H.J.W.,
 RA Pack K., Bell J.A., Williams C.B., Tomlinson I.P.M.;
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
 RT "The APC variants I1307K and E1317Q are associated with colorectal
 RT tumors, but not always with a family history.";
 RL proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [22]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400259; PubMed=9731533;
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
 RT "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews.";
 RL Nat. Genet. 20:62-65(1998).

Query Match Best Local Similarity 99.7%; Score 14533; DB 1; Length 2843; Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	MAAASYDOLLKOVEALKWENSLRQLEDNENHLLTKLETEASNKEVUKQLOGSIEDEAM	60
Db	1	MAAASYDOLLKOVEALKWENSLRQLEDNENHLLTKLETEASNKEVUKQLOGSIEDEAM	60
Qy	61	ASSGOIDILLERKEINLNDSSNPPGVKLSKMSLRYSRGREGSVASRSCECPVPMGSPFR	120
Db	61	ASSGOIDILLERKEINLNDSSNPPGVKLSKMSLRYSRGREGSVASRSCECPVPMGSPFR	120
Qy	121	RGVNGSRSTGYLBELEKERSLLADLDKEEKEXDYIAQONLTGKIDSLPTENPSL	180
Db	121	RGVNGSRSTGYLBELEKERSLLADLDKEEKEXDYIAQONLTGKIDSLPTENPSL	180
Qy	181	QTDLTRQLEYEARQIRVAMEBOLGTCQDMKRAORRARIQOIEKDLRLRQLLOSOAT	240
Db	181	QTDLTRQLEYEARQIRVAMEBOLGTCQDMKRAORRARIQOIEKDLRLRQLLOSOAT	240
Qy	241	EAERSQNHGTGSHDAERQEGVGGEINMATSGQGGTTRMDHETASVLSSSSTHSA	300
Db	241	EAERSQNHGTGSHDAERQEGVGGEINMATSGQGGTTRMDHETASVLSSSSTHSA	300
Qy	301	PRRLTSHLGTKEVYVLSLMLGTHDKDDMSHTLLAMSSQDSCISMROSCCLPLLIQLL	360
Db	301	PRRLTSHLGTKEVYVLSLMLGTHDKDDMSHTLLAMSSQDSCISMROSCCLPLLIQLL	360
Qy	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHISQPDCKRREIRVLHLEQIRAYCETC	420
Db	361	HGNDKDSVLLGNSRGSKEARASAAALHNIHISQPDCKRREIRVLHLEQIRAYCETC	420
Qy	421	WEMQEAHEPGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMELGGLOAIAELLQ	480
Db	421	WEMQEAHEPGMDQKNPMPAPVEHQICPAVCLMKLSFDEEHRHAMELGGLOAIAELLQ	480
Qy	481	VDCVMGLTNDHYSITLRYAGWALTNLTFGVDVANKATLCSMKGCMRALVAQLKSESDL	540
Db	481	VDCVMGLTNDHYSITLRYAGWALTNLTFGVDVANKATLCSMKGCMRALVAQLKSESDL	540
Qy	541	QOVIASVLNLSWRADVNSKKTILREYGVSKALMECALEYKKESTLKSVLNLSAHT	600
Db	541	QOVIASVLNLSWRADVNSKKTILREYGVSKALMECALEYKKESTLKSVLNLSAHT	600
Qy	601	ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGGILRNYSLLIATNEDHROILRENN	660
Db	601	ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGGILRNYSLLIATNEDHROILRENN	660
Qy	661	CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGAVSMKLNLIHSHKMIAM	720
Db	661	CIQTLLOHLKSHSLTIVSNACGTLNLSARNPKQOEALWDMGAVSMKLNLIHSHKMIAM	720
Qy	721	GSAALRNLMANRPAXYKDANIMSPGSSILPSLHVRKQKALEAELDAQHLSSETFDNIDLS	780
Db	721	GSAALRNLMANRPAXYKDANIMSPGSSILPSLHVRKQKALEAELDAQHLSSETFDNIDLS	780
Qy	781	PKASHSKORHKOSLYGVDFDNRHDDNRSDFNFTGNMTVLSPYINTVLTPSSSSSRGS	840
Db	781	PKASHSKORHKOSLYGVDFDNRHDDNRSDFNFTGNMTVLSPYINTVLTPSSSSSRGS	840
Qy	841	LDSSRSSEKORSERERIGLGNYPHATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS	900
Db	841	LDSSRSSEKORSERERIGLGNYPHATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS	900
Qy	901	QEDRSSTTELHCVTDERNALRBSNAATHSNFTKSENRCNCPYAKLEYKRSS	960
Db	901	QEDRSSTTELHCVTDERNALRBSNAATHSNFTKSENRCNCPYAKLEYKRSS	960
Qy	961	NDLSINSSNDGKGKQKPSIESYSEDDSEKFCYGYPADLAHIIHSANHMDNDGE	1020
Db	961	NDLSINSSNDGKGKQKPSIESYSEDDSEKFCYGYPADLAHIIHSANHMDNDGE	1020
Qy	1021	LDTPINYSKYSDQNLNQRQSPQSNRWARPKHIIIEDEIKOSEQRQSNQSTTYPVYTE	1080
Db	1021	LDTPINYSKYSDQNLNQRQSPQSNRWARPKHIIIEDEIKOSEQRQSNQSTTYPVYTE	1080
Qy	1081	STDDKHLKQPHGQOECVSPYBSRGANGSETNRVGSNGHGINCVSOSLCOEDDYDDKP	1140
Db	1081	STDDKHLKQPHGQOECVSPYBSRGANGSETNRVGSNGHGINCVSOSLCOEDDYDDKP	1140
Qy	1141	TNYSERYSEEQHEEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFPSKS	1200
Db	1141	TNYSERYSEEQHEEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFPSKS	1200
Qy	1201	SSQSSKTEHMSSESSENTSTPSSNAKRONCLHPSSSAQSRGQPOKAATCKVSSINQETIQ	1260
Db	1201	SSQSSKTEHMSSESSENTSTPSSNAKRONCLHPSSSAQSRGQPOKAATCKVSSINQETIQ	1260
Qy	1261	TYCVEDTPICFSRCSSLSLSSAEDIGCQNTTQEADSAANTLQIAIEIKGIGTRSAEDPV	1320
Db	1261	TYCVEDTPICFSRCSSLSLSSAEDIGCQNTTQEADSAANTLQIAIEIKGIGTRSAEDPV	1320
Qy	1321	SEVPANVQHPRTKSRLOGSSLSSEBARHKAFFPSGAKSPSKGAQTPKSPPEHYVQET	1380
Db	1321	SEVPANVQHPRTKSRLOGSSLSSEBARHKAFFPSGAKSPSKGAQTPKSPPEHYVQET	1380
Qy	1381	PLMFSRCTSVSSLDSPESRSIASSVQSPGSMVSGIISPSDLPDPSGTMPSPRSKTPP	1440
Db	1381	PLMFSRCTSVSSLDSPESRSIASSVQSPGSMVSGIISPSDLPDPSGTMPSPRSKTPP	1440
Qy	1441	PPQTAQTKREVKNKAPTAEKESGPKQAAVNAQVQVLPDADTLHLHFAESTPDGF	1500
Db	1441	PPQTAQTKREVKNKAPTAEKESGPKQAAVNAQVQVLPDADTLHLHFAESTPDGF	1500
Qy	1501	SCSSLSALSILDEPFIQKQVLELIMPPVQENDNGNETSEOPKESNENQKEAEKTIJSE	1560
Db	1501	SCSSLSALSILDEPFIQKQVLELIMPPVQENDNGNETSEOPKESNENQKEAEKTIJSE	1560
Qy	1561	KDLLDDDDDDDIIELEECIIISAMPTKSSRGKPKPAQTAASKLPPPPVAKPSQLPVYKLLPS	1620
Db	1561	KDLLDDDDDDDIIELEECIIISAMPTKSSRGKPKPAQTAASKLPPPPVAKPSQLPVYKLLPS	1620
Qy	1621	QNRLOPKHVSFTPGDDMPVYCVGEGTPIINFSTATSLSDLTIESPPNELAAGEVGRGAQ	1680
Db	1621	QNRLOPKHVSFTPGDDMPVYCVGEGTPIINFSTATSLSDLTIESPPNELAAGEVGRGAQ	1680
Qy	1681	SBEPEKRTTPTTEGRSTDEAGGKTSVVTIPELDDNKAESGDILAEICINAMPKGSKHP	1740
Db	1681	SBEPEKRTTPTTEGRSTDEAGGKTSVVTIPELDDNKAESGDILAEICINAMPKGSKHP	1740
Qy	1741	PRVKIMQVQOQASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYRTRVRKNADSNNLN	1800
Db	1741	PRVKIMQVQOQASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYRTRVRKNADSNNLN	1800
Qy	1801	AERFSDNKSCKONLNKNSKDFNDKLPNNEDRVGSGFADSPHHTPIEGTPTCFGRND	1860
Db	1801	AERFSDNKSCKONLNKNSKDFNDKLPNNEDRVGSGFADSPHHTPIEGTPTCFGRND	1860
Qy	1861	SLSLDFDDDDVDLSREKAEALRKAENKESAKVTSTHETSTNQQSANKTQAIKQPINR	1920
Db	1861	SLSLDFDDDDVDLSREKAEALRKAENKESAKVTSTHETSTNQQSANKTQAIKQPINR	1920
Qy	1921	QGPPILOKQSTFPQSSKQIPDRCAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDENN	1980
Db	1921	QGPPILOKQSTFPQSSKQIPDRCAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDENN	1980
Qy	1981	NKENPIKETETPPDQSGEPSPKQASGYAPKSFHVHEDTFCFSRNSLSLSLSDIDELLQ	2040
Db	1981	NKENPIKETETPPDQSGEPSPKQASGYAPKSFHVHEDTFCFSRNSLSLSLSDIDELLQ	2040
Qy	2041	ECISSAMPKKKPSRLKQDNKESPRNNGGILGDELTLDKDIQRPDSEHGLSPDSENF	2100
Db	2041	ECISSAMPKKKPSRLKQDNKESPRNNGGILGDELTLDKDIQRPDSEHGLSPDSENF	2100
Qy	2101	WKAIQEGANSIVSSLHQAAAAACLSRQASDSDSLKSGISLGSFPFLTPDQBEKEFT	2160

Db 2101 WKAQEGANSVSSLHQAAAAACLSRQASSSDLSLSKSLGSPFHLTPDGEKPT 2160
 QY 2161 SNKGPRILKPEKSTLETKKIESSEKGIKGGKVKYSLITKGVRSNBEISQMKQPIQAN 2220
 Db 2161 SNKGPRILKPEKSTLETKKIESSEKGIKGGKVKYSLITKGVRSNBEISQMKQPIQAN 2220
 QY 2221 MPSISRGRTMIHICPVNRSSSTSPVSKKGPPLKTPASKSEGCQTATTGPRGAKPVKS 2280
 Db 2221 MPSISRGRTMIHICPVNRSSSTSPVSKKGPPLKTPASKSEGCQTATTGPRGAKPVKS 2280
 QY 2281 ELSPVARTSIOGSSKAPSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Db 2281 ELSPVARTSIOGSSKAPSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
 QY 2341 KLSQLPRTSSPTASTKSSGSKMSYSPGRMSQOQLTKOTGLSKVASSIPRSESASKG 2400
 Db 2341 KLSQLPRTSSPTASTKSSGSKMSYSPGRMSQOQLTKOTGLSKVASSIPRSESASKG 2400
 QY 2401 LNQMNGNGANKVELSRMSSTKSSGESDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 Db 2401 LNQMNGNGANKVELSRMSSTKSSGESDRSERPVLVROSTFIKEAPSPTLRKLKLESA 2460
 QY 2461 SFESLSPSRPASPTRSQAQTPVLSPSLPMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 Db 2461 SFESLSPSRPASPTRSQAQTPVLSPSLPMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 QY 2521 RPAKHDIARSHSPRLPINRGSTWKRHSKSSSLPRVSTWRTGSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSPRLPINRGSTWKRHSKSSSLPRVSTWRTGSSSSILSASSES 2580
 QY 2581 SEKAKSEKHNVSIGTKQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Db 2581 SEKAKSEKHNVSIGTKQSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 QY 2641 KTLIQVAPAVSKTVDVWRLEDPCINPNRSGRPTGNTPPVDSVSEKANPNIKOSKON 2700
 Db 2641 KTLIQVAPAVSKTVDVWRLEDPCINPNRSGRPTGNTPPVDSVSEKANPNIKOSKON 2700
 QY 2701 QAKQNVNGSVPMRTVGLNRLNFIQVDAPDQKTEIKPGQNNPVVSETNESPIVERT 2760
 Db 2701 QAKQNVNGSVPMRTVGLNRLNFIQVDAPDQKTEIKPGQNNPVVSETNESPIVERT 2760
 QY 2761 PFSSSSSKHSPSGTVAARTPPNNYSPRKSADSTSRPQIPTPVNNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARTPPNNYSPRKSADSTSRPQIPTPVNNNTKKRDSKT 2820
 QY 2821 DSTESSGTQSPKRHSGLYVTSV 2843
 Db 2821 DSTESSGTQSPKRHSGLYVTSV 2843

RESULT 2
 APC_RAT ID-APC RAT STANDARD; PRT; 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RT "cDNA cloning of the rat APC gene and assignment to chromosome 18";
 PL Mamm. Genome 6:746-748(1995).

Query Match 90.3%; Score 13168.5; DB 1; Length 2842;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 2557; Conservative 123; Mismatches 158; Indels 9; Gaps 8;

QY 1 MAASYYDQLLKQVALKMNSLNKQLEEDNSNHLTKLETASNKKEVLKOLQGSIEDAM 60
 Db 1 MAASYYDQLLKQVALKMNSLNKQLEEDNSNHLTKLETASNKKEVLKOLQGSIEDAM 60
 QY 61 ASSGQIDLLERLKEINLDSSNFPVGVKLSKMSLSYSGSRGSSVSRSGECSVPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEFNLD-SNFPVGVKLSKMSLSYSGSRGSSVSRSGECSVPVPMGSPFR 118
 QY 121 RGFVNGSPESGYLEELEKERSILLADLKKEEKDNYAQLQNLTKRIDSPLTENFSL 180
 Db 119 RAFVNGSPESGYLEELEKERSILLADLKKEEKDNYAQLQNLTKRIDSPLTENFSL 178
 QY 181 QTDLTRQLEYEARQIRVAMEEQGTQDMKRAQRRIARIQTEKIDILRIQLQSOAT 240
 Db 179 QTDLTRQLEYEARQIRVAMEEQGTQDMKRAQRRIARIQTEKIDILRIQLQSOAA 238
 QY 241 BAERSSQKHETGSHDAERQNEGCGVEINWATSGNGCGSTTRMDHETASVLSSSSTHSA 300

RN [2]
 RP MUTAGENESIS.
 RC STRAIN=Sprague-Dawley, and Fischer 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914 (1995).
 CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 and participates in Wnt signaling. APC activity is correlated with
 its phosphorylation state (By similarity).
 CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
 axin (By similarity).
 CC -!- PTM: Phosphorylated by GSK3B (By similarity).
 CC -!- SIMILARITY: Contains 7 ARM repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D38629; BAA07609.1; -.
 DR HSP; Q02248; 3BCT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 4.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS0176; ARM_REPEAT; 1.
 KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
 KW Repeat.
 FT DOMAIN 1 728 LEU-RICH.
 FT DOMAIN 1 62 COILED COIL (POTENTIAL).
 FT DOMAIN 125 260 COILED COIL (POTENTIAL).
 FT REPEAT 451 493 ARM 1.
 FT REPEAT 503 545 ARM 2.
 FT REPEAT 546 589 ARM 3.
 FT REPEAT 590 636 ARM 4.
 FT REPEAT 637 681 ARM 5.
 FT REPEAT 682 723 ARM 6.
 FT REPEAT 724 765 ARM 7.
 FT DOMAIN 739 2831 SER-RICH.
 FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 HIGHLY CHARGED.
 FT MUTAGEN C->R; IN AN IQ-INDUCED COLON TUMOR.
 SQ SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E8F47 CRC64;

Db 239 EAERSSQSKHETAHEARQLEGQGVAEISNLIATSGSQSAAARVDHETAGVLSSGTHSA 238
Qy 301 PRRLTSHLGTKEVYISLLSMGLTHDKDMSRTLLAVSSQSDSCISMQRGCLPLLIQLL 360
Db 299 PRRLTSHLGTKEVYISLLSMGLTHDKDMSRTLLAVSSQSDSCISMQRGCLPLLIQLL 358
Qy 361 HGNDKDSVLLGNSRGSKARASAAALNIIHSDPDDKRGREIRVHLLEQIYAYCETC 420
Db 359 HGNDKDSVLLGNSRGSKARASAAALNIIHSDPDDKRGREIRVHLLEQIYAYCETC 418
Qy 421 WEMQEAHEGMDQDNMPAIVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIELLQ 480
Db 419 WEMQEAHEGMDQDNMPAIVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAIELLQ 478
Qy 481 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSKMGCMRALVAQLKSEEDL 540
Db 479 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSKMGCMRALVAQLKSEEDL 538
Qy 541 QCVIASVLRLNLSRADVNSKTLREVGSVKALMECALVEKESITLKSVLKALMNLASHCT 600
Db 539 QCVIASVLRLNLSRADVNSKTLREVGSVKALMECALVEKESITLKSVLKALMNLASHCT 598
Qy 601 ENKADICAVDGALAPLVGTLTYRGTNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 660
Db 599 ENKADICAVDGALAPLVGTLTYRGTNTLAIIESGGGILRVNSVSLIATNEDHROILRENN 658
Qy 661 CLQTLLOHLKSHSLTIVSNACGTLNLSARPKDOEALMDMGAVSMKLNLIHSHKXIAM 720
Db 659 CLQTLLOHLKSHSLTIVSNACGTLNLSARPKDOEALMDMGAVSMKLNLIHSHKXIAM 718
Qy 721 GSAAALRLNLMANRPKAYKDANIMSPGSLPSLHVVRKQKALEAELDAQHLSLTFDNIDLS 780
Db 719 GSAAALRLNLMANRPKAYKDANIMSPGSLPSLHVVRKQKALEAELDAQHLSLTFDNIDLS 778
Qy 781 PKASHRSQKHQSLYGDVFTNDRDNRDNFTGNMTVLSPLYNTTIVLPSSSSRG 840
Db 779 PKASHRSQKHQSLYGDVFTNDRDNRDNFTGNMTVLSPLYNTTIVLPSSSSRG 838
Qy 841 LDSSSEKDRSLERERG:GLGNHYHATENPGTSSKRGILQISTTAAQIAKWEEVSALHTS 900
Db 839 LDSSSEKDRSLERERG:GLSYHATENPGTSSKRGILQISATAAQIAKWEEVSALHTS 898
Qy 901 QEDRSSTTELHCVTDERNALRRSAATHNTVNTFKSENSNPTCMYPKALEYKRSS 960
Db 899 QEDRSSTTELHCVTDERNALRRSAATHNTVNTFKSENSNPTCMYPKALEYKRSS 958
Qy 961 NDLSNVSNDGYSKRGQKPSIYSYEDDESKFCYGOYPADLAHAKHSANHMDNDGE 1020
Db 959 NDLSNVSNDGYSKRGQKPSIYSYEDDESKFCYGOYPADLAHAKHSANHMDNDGE 1018
Qy 1021 LPTPINYSILKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQSRNQSTTVPVYTE 1080
Db 1019 LPTPINYSILKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQSRNQSTTVPVYTE 1078
Qy 1081 STDDXHLKFPQHFQOECVSPYRSGANGSEFTNRVGNHGINQVYSQICQEDDYEDDKP 1140
Db 1079 NTDDXHLKFPQHFQOECVSPYRSGANGSEFTNRVGNHGINQVYSQICQEDDYEDDKP 1138
Qy 1141 TNYSEYSEEEHEEERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKSFSSKS 1200
Db 1139 TNYSEYSEEEHEEERPTNYSIKYNEKRVHDQPIDYSLKYATDIPSSQKSFSSKS 1198
Qy 1201 SSCQSKSTHEMSSSENTSTPSNAKRNQOLHPSSAQSRSGOPQAACTCKVSSINQETIQ 1260
Db 1199 PSVQGTKEHNSPSSAASAPSNKROSLHPSSAQ-RNGQTPKGTACKVPSINQETMQ 1257
Qy 1261 TYCVEDTPICFRCSLSLSAEDRIGCNQTTQEADSAANTLQIAEIKGIGTRGAEDPV 1320
Db 1258 TYCVEDTPICFRCSLSLSAEDRIGCNQTTQEADSAANTLQIAEIKGIGTRGAEDPV 1317
Qy 1321 SEYPAYSQHPRTKSSQLQSSSLSSESARHAKVEFPSSGAKSPSKGAQTTPKSPPEHYOET 1380
Db 1318 SDVPAYSQHPRTKSSQLQSSSLSSESARHAKVEFPSSGAKSPSKGAQTTPKSPPEHYOET 1377

Qy 1381 PLMFSECTSVSLDSFESRSIASSVQSEPCSGMWGSIISPSDLPSPGOTMPPSSSKT-- 1438
Db 1378 PLMFSECTSVSLDSFESRSIASSVQSEPCSGMWGSIISPSDLPSPGOTMPPSSSKT-- 1437
Qy 1439 PPPPPPTAGTKKEVPKQKAPTAEKRESGPKQAANAAVQVQVLPDADTLLHFATESTPD 1498
Db 1438 PPPPPPTAGTKKEVPKQKAPTAEKRESGPKQAANAAVQVQVLPDADTLLHFATESTPD 1497
Qy 1499 GFSCSSSLGALSJDEPFIQKVELRIMPVQENDNGNETESEPQKESNENQEKAKETD 1558
Db 1498 GFSCSSSLGALSJDEPFIQKVELRIMPVQENDNGNETESEPQKESNENQEKAKETD 1556
Qy 1559 SEKDLDDDDDDIEILEECIISAMPTKSSRRKPKQAQASKLPPPVARKPSOLPVYKLL 1618
Db 1557 SEKDLDDDDDDIEILEECIISAMPTKSSRRKPKQAQASKLPPPVARKPSOLPVYKLL 1616
Qy 1619 PSQNRLOPQKHVSFTFGDDMPRVYCVGTEPIINFSTATISLDTIESPPNELAAGEVRGG 1678
Db 1617 PSQNRLOPQKHVSFTFGDDMPRVYCVGTEPIINFSTATISLDTIESPPNELAAGEVRGG 1676
Qy 1679 AOSGEFEKDDTIPTEGRSTDEAGGKTSVTIPELDNKAEEGDIABEININSAMPKXSH 1738
Db 1677 AOSGEFEKDDTIPTEGRSTDEAGGKTSVTIPELDNKAEEGDIABEININSAMPKXSH 1736
Qy 1739 KPRVRVKIMDVQQAASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKVN 1798
Db 1737 KPRVRVKIMDVQQAASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKVN 1796
Qy 1799 LNAERFVSNKSKQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHYPIEGTPTCESR 1858
Db 1797 VNTETPSDNKSKKQKQNLKNNKDFNDKLPNNEDVRGSAFDSPHYPIEGTPTCESR 1856
Qy 1859 NDLSLSDLDLDDDDVLSREKAEIRKAKENKESAKVTSHTLTSTNQSSANKTQAIKQPI 1918
Db 1857 NDLSLSDLDLDDDDVLSREKAEIRKAKENKESAKVTSHTLTSTNQSSANKTQAIKQPI 1916
Qy 1919 NRQOPKILQKQTFPQSSKDIIDRGAATDEKQNFATENTPVCFSHNSLSLSIDQEE 1978
Db 1917 NRQOPKILQKQTFPQSSKDIIDRGAATDEKQNFATENTPVCFSHNSLSLSIDQEE 1976
Qy 1979 -NNKENEPKETEPDPSQSEKPSKPOASGVAPKSFVEDTPVCFNSNSLSLSIDSEDD 2037
Db 1977 -NNKENEPKETEPDPSQSEKPSKPOASGVAPKSFVEDTPVCFNSNSLSLSIDSEDD 2036
Qy 2038 LLOECISSAMPKRRPSRLKGEWQSPKRVGSLAEDLTDLKDIQRPSEHGLSPDSE 2096
Db 2037 LLOECISSAMPKRRPSRLKGEWQSPKRVGSLAEDLTDLKDIQRPSEHGLSPDSE 2096
Qy 2098 NFDWKATQEGANSTVSSLIHQ-AAAAACLSQASDSDSILSLKSGISLGSPEHLTPDQEE 2156
Db 2097 NFDWKATQEGANSTVSSLIHQ-AAAAACLSQASDSDSILSLKSGISLGSPEHLTPDQEE 2156
Qy 2157 KPFTSNKGPRIILKPGKSTLTETKKIESKGIKGGKVKYKSLITGKVRNSEISGQMKQP 2216
Db 2157 KPFTSNKGPRIILKPGKSTLTETKKIESKGIKGGKVKYKSLITGKVRNSEISGQMKQP 2216
Qy 2217 LOANPISIRGRWTMIHFGVNRSSSTSPVSKGPPKLTASKSPSEGOATTPSPGAKP 2276
Db 2217 LOANPISIRGRWTMIHFGVNRSSSTSPVSKGPPKLTASKSPSEGOATTPSPGAKP 2276
Qy 2277 SVKSELSPVARQTOIGGSKAPSRSGRSTPSRPAQQLPSRPIQSPGRNSISPGNGI 2336
Db 2277 SVKSELSPVARQTOIGGSKAPSRSGRSTPSRPAQQLPSRPIQSPGRNSISPGNGI 2336
Qy 2337 SPPNKLSQLPRTSSPSTASTKSGSGKRWYTSPPGRQMSQONLTKQTLGSKNASSIPRES 2396
Db 2337 SPPNKLSQLPRTSSPSTASTKSGSGKRWYTSPPGRQMSQONLTKQTLGSKNASSIPRES 2396
Qy 2397 ASKGLNOMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKL 2456
Db 2397 ASKGLNOMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKL 2456

```

QY 2457 EESASFESLSPSSRPASPTRSQAOPTVLPSPSLPMSLSTHSSVOAGGWRKLPNPLSPSTIE 2516
DB 2457 EESASFESLSPSSRPDSPTRSQAOTPVLPSPSLPMSLSTHPSVOAGGWRKLPNPLSPSTIE 2516
QY 2517 YNDGRPAKHDIARSHGESPRLPINSRGTWKRSHKSSSLPRVSTWRTGSSSSILSA 2576
DB 2517 YSDGRPSKXHDIAHSHGESPRLPVNAGTWKRSHKSSSLPRVSTWRTGSSSSILSA 2576
QY 2577 SSESSEKAKSDEKXVNSISGTOKSKENQVSAKGTWRKIKENEFSPNTSTQTVSSGATN 2636
DB 2577 SSESSEKAKSDEKXVNSVPGPRQKENVQVTKGTWRKIKESISPTNTVSTQTVSSGAAS 2636
QY 2637 GAESKTLIYQAPAVSKTQVWRIEDCPINNPRSGSPGTGNTPPVIDSSEKANPNIKD 2696
DB 2637 GAESKTLIYQAPAVSKTQVWRIEDCPINNPRSGSPGTGNTPPVIDSSEKANPNIKD 2696
QY 2697 SKDQAQKQNVGSGVPMRTVGLNRLTSFIQVADPDKQTEIKPGQNNPVVSETNESPI 2756
DB 2697 SKDTQKQSVGSGS- PVQTVGLNRLNSFIQVEAPEQKGTETKAGQSGPAPVATGETCM 2755
QY 2757 VERTPFSSSSSKHSSSGTVAARVTFFNPNPSPRKSADSTGAPSGQIPTPVNNNTKKR 2816
DB 2756 AERTPFSSSSSKHSSSGTVAARVTFFNPNPSPRKSADSTGAPSGQIPTPVGSSSTKKR 2815
QY 2817 DSKTDSSTESSGTQSPKHSGLSYLTVSV 2843
DB 2816 DSKTDSSTESSGAQSPKHSGLSYLTVSV 2842

RESULT 3
ID APC MOUSE
AC Q61315, Q62044; STANDARD; PRT: 2845 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein) (mAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RP STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RA "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene.";
RL Science 256:668-670(1992).
RN [2]
ERRATUM.
RP Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114(1992).
RN [3]
SEQUENCE OF 1-45 FROM N.A.
RP STRAIN=BALB/c; TISSUE=Liver;
RA Dicker F., Lambertz S., Reitmair A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RL region segments.";
RN Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
ALTERNATIVE SPLICING.
RP MEDLINE=94061824; PubMed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RA "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL Cancer Res. 53:5589-5591(1993).
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
CC and participates in Wnt signaling. APC activity is correlated with
CC its phosphorylation state (By similarity).
CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds

```

```

CC axin (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q61315-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q61315-2; Sequence=VSP_0041116;
CC Name=3;
CC IsoId=Q61315-3; Sequence=VSP_0041117;
CC Name=4;
CC IsoId=Q61315-4; Sequence=VSP_0041116, VSP_0041117;
CC -!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
CC lung, brain, stomach, intestine, testis and ovary.
CC -!- PTM: Phosphorylated by GSK3B (By similarity).
CC -!- SIMILARITY: Contains 7 ARM repeats.
CC -----
CC This SWISS-PROT entry is copy-right. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88127; AAB59632.1; -.
CC EMBL; U02937; AAB03443.1; -.
CC FIR; I49505; I49505.
CC HSP; Q02248; 3BCT.
CC MGD; MGI:88039; Apc.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0008013; F:beta-catenin binding; IDA.
CC GO; GO:0009982; P:anterior/posterior pattern formation; IMP.
CC GO; GO:0009799; P:axis specification; IMP.
CC GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
CC GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR000225; Armadillo.
CC Pfam; PF00514; Armadillo_seg. 4.
CC SMART; SM00185; ARM_5_REPEAT; 1.
CC PROSITE; PS0176; ARM_REPEAT; 1.
CC Wnt signaling pathway; Anti-oncogene; Phosphorylation;
KW Alternative splicing; Repeat; Coiled coil.
FT DOMAIN 1 61
FT DOMAIN 125 245
FT DOMAIN 1 728
FT REPEAT 451 493
FT REPEAT 503 545
FT REPEAT 546 589
FT REPEAT 590 636
FT REPEAT 637 681
FT REPEAT 682 723
FT REPEAT 724 765
FT REPEAT 739 783
FT DOMAIN 1130 1156
FT DOMAIN 1556 1575
FT DOMAIN 1864 1891
FT VARSPLIC 243 276
FT VARSPLIC 310 410
FT VARIANT 120 120
FT VARIANT 493 493
FT VARIANT 797 797
FT VARIANT 1330 1330
FT VARIANT 1618 1618
FT VARIANT 2294 2294
FT VARIANT 2496 2496
FT VARIANT 2523 2523
FT VARIANT 2813 2813
FT VARSPLIC 2845 AA; 311086 NW; 145CA73CF570A499 CRC64;
SQ SEQUENCE 2845 AA; 311086 NW; 145CA73CF570A499 CRC64;
Query Match 90.2%; Score 13148; DB 1; Length 2845;

```

	Best Local Similarity	90.0%;	Pred. No. 0;	
	Matches 2566;	Conservative 113;	Mismatches 159;	Indels 12; Gaps 10;
Qy	1	MAAASYDQLLKQVEALKWENSLRQELDEDNSNHLTKLETEASNKKVYLKQLOGSIEDAM	60	
Db	1	MAAASYDQLLKQVEALKWENSLRQELDEDNSNHLTKLETEASNKKVYLKQLOGSIEDTM	60	
Qy	61	ASSGQIDLLRLKELNLDSDNFPQVKLRKMSLRASYSGREGSVSSRGEGCSPVPMGSPFR	120	
Db	61	-TSGQIDLLRLKELNLD-SNFPQVKLRKMSLRASYSGREGSVSSRGEGCSPVPMGSPFR	118	
Qy	121	RGFVNGSREGTGYLEELKERSLLADLDKEEKKDWYAAQLQNLTKRIDSLPLTENFSL	180	
Db	119	RTFVNGSREGTGYLEELKERSLLADLDKEEKKDWYAAQLQNLTKRIDSLPLTENFSL	178	
Qy	181	QDILTRQLEYEARQIRVAMEEQIGTQDQMEKRAQRRIARIQIQIEKQILRIRQLLOSOAT	240	
Db	179	QDITRQLEYEARQIRAAVEEQIGTQDQMEKRAQRRIARIQIQIEKQILRVRQLLOSOAA	238	
Qy	241	EARRSSQNHETGSHDAERQNEGQGVGEIINMATSGNGGSGTTRMDHETASVLSSSSTHSA	300	
Db	239	EARRSSQRHDAASHAGRQHEGHGVAESNTAASSGQSPATRVVDHETASVLSSTGTHSA	298	
Qy	301	PRRLTSHLGTKVEMVYLLSLVGLTHDKDDMSRTLLAMSSQDSCISMROGCPLPLLQLL	360	
Db	299	PRRLTSHLGTKVEMVYLLSLVGLTHDKDDMSRTLLAMSSQDSCISMROGCPLPLLQLL	358	
Qy	361	HGNDKDSVILGNSRGSKEARASAAAHNIIHSQPDQKGRREIRVLHLLEQIRAYCETC	420	
Db	359	HGNDKDSVILGNSRGSKEARASAAAHNIIHSQPDQKGRREIRVLHLLEQIRAYCETC	418	
Qy	421	WENQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDBEHRHANNELGLOAIELLQ	480	
Db	419	WENQEAHEQMDQKNPMPAPVEHQICPAVCVLMKLSFDBEHRHANNELGLOAIELLQ	478	
Qy	481	VDCEMYGLTNDHYISITIRVAGMALTNLTGQDVANKATLCMSKQCMRALVAQLKSESED	540	
Db	479	VDCEMYGLTNDHYISITIRVAGMALTNLTGQDVANKATLCMSKQCMRALVAQLKSESED	538	
Qy	541	QQVIAVLRLNLSRADVNSKKTLEVGSVKALMECALVEKKESTLKVLSALWNLASHCT	600	
Db	539	QQVIAVLRLNLSRADVNSKKTLEVGSVKALMECALVEKKESTLKVLSALWNLASHCT	598	
Qy	601	ENKADIICAVDQALAFVLTITLYRSQTNLTALIESGGGILRVNSSLIATNEDHROI	660	
Db	599	ENKADIICAVDQALAFVLTITLYRSQTNLTALIESGGGILRVNSSLIATNEDHROI	658	
Qy	661	CLQTLLOHLKSHSLTIVSNAGTLWNLARNPKQDEALWDMGVSMKLNLIHSHKXMIAM	720	
Db	659	CLQTLLOHLKSHSLTIVSNAGTLWNLARNPKQDEALWDMGVSMKLNLIHSHKXMIAM	718	
Qy	721	GSAAALRLNLMANRPAPKYKDNIMSPGSSLPSPHVRKQKALEAEALDAOHLSETFDN	780	
Db	719	GSAAALRLNLMANRPAPKYKDNIMSPGSSLPSPHVRKQKALEAEALDAOHLSETFDN	778	
Qy	781	PKASHRSKQHKQSLGQYVFDTRHDDNSDNFTGNTVLSPLYLNTTTLVPSSSSRG	840	
Db	779	PKASHRSKQHKQSLGQYVFDTRHDDNSDNFTGNTVLSPLYLNTTTLVPSSSSRG	838	
Qy	841	LDSSRSKDRSLERERGIGLGNHYHPATENPGCTSSKRGLOISTTAAQIAKWMEEVSA	900	
Db	839	LDSSRSKDRSLERERGIGLSAYHPTTENAGTSSKRGLOITTTTAAQIAKWMEEVSA	898	
Qy	901	QEDRSSGSTTELHCVTDERNALRASSAAHTHSNTYNTFTKSENSNRTCSMPYAKLEY	960	
Db	899	QEDRSSASTTEFHCVADRSARASSASHTSNTYNTFTKSENSNRTCSMPYAKLEY	958	
Qy	961	NDLSNYSVNSNDGYKRGQMKPSIESYEDDSKFCSGYQYVADLAHKIHSANHHDDN	1020	
Db	959	NDLSNYSVNSNDGYKRGQMKPSIESYEDDSKFCSGYQYVADLAHKIHSANHHDDN	1018	
Qy	1021	LDPTINTYLSKYDDELNSGROSQNERWAPKPHIIEDEIKQESQORSQNSQITTPYV	1080	

Db 2319 TKPSPPLSSGD--HEANGFGTSLNPPGFVTTATAKEBAEPHAWERGSWPEGHERSS 2376
 QY 2621 SPTN--STQTVSSGATNGAESKTLIYQM:-----APAVSKT 2654
 Db 2377 RPTDLSSEQLPRKSGSGPPCSLSSEVAGPQGCATDPRPHCGBELSPSFLNPLPPST 2436
 QY 2655 EDVWVRLEDCEI-----NPPSRGR-SPTGNTPTPVIVDSVSEKANPNIKDSKDNQAKQNVNG 2709
 Db 2437 DSDLSSTEERLACKGRRVVRGPGATGGCPWAD---ETPPTSADSDSSGSSDSDVPE 2493
 QY 2710 SVMPTVGLNRLTS-----FIQVD-APDQKGT-EIKPGON-NEVPVSETNESP-----IV 2757
 Db 2494 TESCPSITAEALDSDGDFLPVDRAGGVSGTHPRPGHDPPTLPDRPSPPRPDVC 2553
 QY 2758 ERTPFSSSSSS-----KHSSFSGTVAARVTFFN-----YNPGRKSSADSTS- 2799
 Db 2554 MADPEGLSSESGRVERLREKRGRRAPGRAPKAPSPARRLDIRGKRSPTPGKGPVDRTSR 2613
 QY 2800 --ARPSQIPTPVANNKK 2815
 Db 2614 TVPRPRSTPQVISAEEK 2631

RESULT 5
 SPEN DROME STANDARD; PRT: 5560 AA.
 AC Q8SX83: Q9NHN1; Q9NJ17; Q9U6C3; Q9VPL1; Q9VPL2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Split ends protein.
 GN SPEN OR CG18497.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), FUNCTION, ALTERNATIVE PROMOTER
 RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS E9 AND
 RP D57
 RP TISSUE=Embryo;
 RX MEDLINE=20025936; PubMed=10556082;
 RA Welllette E.H., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
 RA McGinnis W.;
 RT "spen encodes an RNP motif protein that interacts with Hox pathways
 RT to repress the development of head-like sclerites in the Drosophila
 RT trunk";
 RL Development 126:5373-5385 (1999).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RP TISSUE=Embryo;
 RX MEDLINE=20157049; PubMed=10655223;
 RA Rebay I., Chen F., Hsiao P., Kolodziej P.A., Kuang B.H., Laverty T.,
 RA Suh C., Voas M., Williams A., Rubin G.M.;
 RT "A genetic screen for novel components of the Ras/mitogen-activated
 RT protein kinase signaling pathway that interact with the yan gene of
 RT Drosophila identifies split ends, a new RNA recognition motif-
 RT containing protein";
 RL Genetics 154:695-712 (2000).
 [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Embryo;
 RX MEDLINE=20171275; PubMed=10704397;
 RA Kuang B.H., Wu S.C.-Y., Shin Y.-A., Luo L., Kolodziej P.A.;
 RT "split ends encodes large nuclear proteins that regulate neuronal
 RT cell fate and axon extension in the Drosophila embryo";
 RL Development 127:1517-1529 (2000).
 [4]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=Berkeley;

EX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavalier S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 [5]
 RN REVISIONS, AND ALTERNATIVE SPLICING.
 RP MEDLINE=22426069; PubMed=12537572;
 RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kunkler J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.D., Whitfield E.J., Bayraktaroglu L., Serman B.P.,
 RA Bettencourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 [6]
 RP SEQUENCE OF 424-2002 FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Embryo;
 EX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celinker S.E.;
 RT "A Drosophila full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 [7]
 RN FUNCTION.
 RP MEDLINE=20253107; PubMed=10790398;
 RX Lane M.E., Elend M., Heidmann D., Herr A., Marzodko S., Herzog A.,
 RA Lehner C.F.;
 RT "A screen for modifiers of cyclin E function in Drosophila
 RT melanogaster identifies Cdk2 mutations, revealing the insignificance
 RT of putative phosphorylation sites in Cdk2";
 RL Genetics 155:233-244 (2000).
 [8]
 RN FUNCTION ON EGF RECEPTOR PATHWAY.
 RP MEDLINE=20414403; PubMed=10959845;

RA Chen F., Rebay I.,
RT "split ends, a new component of the Drosophila EGF receptor pathway,
RT regulates development of midline glial cells.",
RL Curr. Biol. 10:943-946(2000).
RN [9]
RP FUNCTION ON WG PATHWAY.
RX MEDLINE=22668876; PubMed=12783785;
RA Lin H.V., Dorquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.,
RT "Splits ends is a tissue/promoter specific regulator of Wingless
RT signaling",
RL Development 130:3125-3135(2003).
CC -!- FUNCTION: Probable corepressor protein, which regulates different
CC key pathways such as the EGF receptor and Wg pathways. Involved in
CC neuronal cell fate, survival and axon guidance, cell cycle
CC regulation and repression of head identity in the embryonic trunk.
CC May act with the Hox gene Deformed and the EGF receptor signaling
CC pathway. Positive regulator of the Wg pathway in larval tissues
CC but not in embryonic tissues. May act as a transcriptional
CC corepressor protein, which repress transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative promoter;
CC Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
CC alternative promoters;
CC Event-Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q8SX83-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8SX83-2; Sequence=VSP_008565, VSP_008566;
CC Note=No experimental confirmation available;
CC Name=3; Synonyms=Spentl;
CC IsoId=Q8SX83-3; Sequence=VSP_008567;
CC Note=Produced by alternative splicing of isoform 1;
CC Name=4; Synonyms=Spens;
CC IsoId=Q8SX83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
CC Note=Produced by alternative splicing of isoform 2;
CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
CC in stage 3 embryos, and in blastoderm cells, including pole cells.
CC Expressed throughout the rest of embryogenesis. Later, it is
CC expressed at higher level in epidermal cells and CNS.
CC -!- DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
CC zygotically.
CC -!- SIMILARITY: Belongs to the Spn family.
CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
CC -!- SIMILARITY: Contains 1 SPOC domain.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF138205; AAF13218.1; -;
DR EMBL; AF138462; AAF26299.1; -;
DR EMBL; AF221715; AAF34661.1; ALT_INIT.
DR EMBL; AE003590; AAF1534.2; -;
DR EMBL; AE003590; AAF1535.2; -;
DR EMBL; AE003590; AAN10511.1; -;
DR EMBL; AY094788; AAM11141.1; ALT_SEQ.
DR HSP; P09651; 1HA1.
DR Flybase; Fgn001697; spen.
DR GO; GO:0007411; P:axon guidance; IMP.
DR GO; GO:008347; P:glia cell migration; IMP.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS00102; RRM_3.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.

DR PROSITE; PS00917; SPOC; 1.
KW transcription regulation; Repressor; Developmental protein;
KW Nuclear protein; Repeat; RNA-binding; Coiled coil;
KW Alternative promoter usage; Alternative splicing.
FT DOMAIN 554 632 RNA-BINDING (RRM) 1.
FT DOMAIN 656 730 RNA-BINDING (RRM) 2.
FT DOMAIN 734 806 RNA-BINDING (RRM) 3.
Query Match 3.3%; Score 474.5; DB 1; Length 5560;
Best Local Similarity 18.2%; Pred. No. 6.5e-10;
Matches 640; Conservative 498; Mismatches 1307; Indels 1065; Gaps 149;
QY 2 AAASYDQLLKQVE---ALKMNSNLQLEDNSNHLTKLETEASNMKEVLKQGGIEDE 58
DB 1786 SLSLSDTTSASVETSASISSSTNGNSLTSAAIHVQKQSQSTFVEEHTKSGTSS 1945
QY 59 AMASSQIDLLERKELNLDNSNFPQVKLRKMSLRYSYSGRSVSRSGEGSPVPMGSP 118
DB 1846 SSSSSKKI-----SSTHD--KLHSHNNRSEDKKIKGDKGNASSDKRKNSS 1891
QY 119 PRGFVNGSRESTGYLEE-----LEKERSILLADLQEEK 153
DB 1892 TSQ-----SSKATPRIEDSSSEADDTADKAENQREKEKREKDLRKQVREK 1947
QY 154 EKDWYVAQNLTKRIDSPLTENFSLQTLTRRQLEYEARQIRVAMEBQLGTCQDMEXR 213
DB 1948 DRK---AQEEREK-----EDRKAKEEREREKKAQEDREKKEBERELREKEOR 1996
QY 214 -----AQRIARIQIEKD---ILRIQLLOSQATEAERSSONKH-ETGSHDAER 259
DB 1997 DKEQKEIREKDLREKQORNRNREKLRDLREKREKREKREKELHREKQDQREH 2056
QY 260 QNEGQGVGEINMATSGNGQSTTRMDHETASVLSSTSTSHAPRLTSHLGTKEVM--VY 316
DB 2057 REKEQSRRAVDVEQGRG---GRM-----RELSSYQSKMDIAGEAS 2095
QY 317 SLTSLMLGTHDKD-----DMSRTLLAMS-----SSQDS 343
DB 2096 SLTAIDQHNKENAMDTIAQGTGASPTSPDNTPKERSKRLSRNPFVRLHRLKRLSSQES 2155
QY 344 CISMROSGCLPLLIQLLHGN-----KDSVLLGNRSGKKEAR-----380
DB 2156 NBSAGGGSCGSGSHQIHEDYVKRIMENSQNI SVHSSNQRLNDRDSEKHSKSFED 2215
QY 381 -----ARASAAHLNIIHSQPDCKGRREIR-----VLHLLEQIRAYCE 418
DB 2216 KNSSSHISIRPHGCGSSASSSKHHRRDKHQKGSASSTIETNSSIEVWVDPISQTKHNLN 2275
QY 419 TCWEMQEAHEPGMDQKNPMPAPVEHOICPAVCVLMKLSFDEHRHAMNELGSLQALAE 478
DB 2276 TSEELQSHQPKKEKER-----EHPSSHANSSSRHKSKKDDHHHREK--KRHSVAES 2326
QY 479 LQVDCMYGLT-NDHYSITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSES 537
DB 2327 TNDTEHTPQCHNPHRRIS-----AAGSGAGELSSAATNTS-----2363
QY 538 EDLQQVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSVALNLSA 597
DB 2364 -----SGKLHQQHRRSVKRSR---GS-----DEGHSSSKSLAKMLLS 2404
QY 598 HCTENKADICAVDGAFLVGLTYRSQTNTLAIIESGGGILRNVSLSIATNEDHRIQLR 657
DB 2405 -----ADSDDTDDA-----2413
QY 658 ENNCLOTLLQHLKSHSLTIVSNACGLTNLSARNPKDQALWDMGAVSLMKNL-----710
DB 2414 -----SKHISIFDIPDDCPNV-----SMYDKVKARSCKNQMQRAEBK 2450
QY 711 -IHSKHMTAMGSAALRNLM--ANRPAYKDIANIMSPGSSLSPLSHVR-----755
DB 2451 KIKAKFSQLKQSRKKKRTSYDGDSDTEFEHQHNSGSS--SFAGRYPGLSSSSDDDD 2508
QY 756 ---KQKALEAEALDAOHLSETFDNIDNLSPKASHRSKQRH-----K 792

QY 2533 SESPSRLPINSGTWKREHSHSSSL-----PRVSTWRR-----TCSSS 2571
DB 4538 LPKGAQTPRESG--RNAQAKTDAVQIINAVGPRRSKORKTTGEQTANLIEBVTASNA 4595
QY 2572 SILGASSSESSKAKSE-----DKHVNISGCT----- 2598
DB 4596 TVAASHLAPPEGAGVESHPQLDRAKEVPSVSVVTPISTAPVSVAAAPVTVFPVPMVFKP 4655
QY 2599 ---KQSKENQVSAG--GTRWIKIKENEF--PTNSTQTVSSGATNG---AESKTLIYQM 2647
DB 4656 TMPQHPKKKATAAAIESYQAINISISGGLPMHTAAPTCKITGGVADAVSKALVDPV 4715
QY 2648 APAVS-----KTEDEVVRIEDCPINPRSRSTGNTPPVIDSVSKANPNIKSDKNQ 2701
DB 4716 TGVITAGMPOGKEGNLPAATAAPANS--SNEDGQAAPPPOLHQQQOQHP--QOOPQOQ 4771
QY 2702 AKQNVGNSVPMRTVGLNRLTSF---IOVDAPQOKGTEIKPGONNPPV-SETNESPIV 2757
DB 4772 ANLQINTYLIFS---GLENPITALGKSVQJB-----TSAAALNKPVSVLVKGNASQVI 4822
QY 2758 ERTFSSSSSKHSPSGTVAARVTP--FNTNPSR-KSSADTSARPSQ-IPTFPVNNNT 2813
DB 4823 QQ-----QOPO-IVAPAKQPIILQONPLPTVLHHAQHTTVRPPQPLKAHVLNRE 4870
QY 2814 KKRDSKTDSTESSGTQSPKR--HSGSYLVT 2841
DB 4871 KNIQOQLPTTKQAVAPQPHSGHMLLT 4900
RESULT 6
MINT HUMAN
ID MINT HUMAN STANDARD; PRT; 3664 AA.
AC Q9G158; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIA0929.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
RP RAR AND MTAL1.
RC TISSUE=Liver, and Pituitary;
RX MEDLINE=21231190; PubMed=11331609;
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
RA Hon M., Evans R.M.;
RT "Sharp, an inducible cofactor that integrates nuclear receptor
RT repression and activation";
RL Genes Dev. 15:1140-1151(2001).
[2]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 294-3664 FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC TISSUE=Embryo, and Teratocarcinoma;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Ito Y., Kawai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]

SEQUENCE OF 2002-3664 FROM N.A.
TISSUE=Brain;
MEDLINE=99246063; PubMed=10231032;
Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hiroseawa M.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
[6]
RP INTERACTION WITH PP4.
RX MEDLINE=21874127; PubMed=11867749;
RA Shi Y., Hon M., Evans R.M.;
RT "The peroxisome proliferator-activated receptor delta, an integrator
RT of transcriptional repression and nuclear receptor signaling";
RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
[7]
RP FUNCTION, AND INTERACTION WITH RBP5UH.
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Koseczka U., Astrantantseff K., Bourteele S., Dillinger K.,
Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway";
RL EMBO J. 21:5417-5426(2002).
[8]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
RX MEDLINE=22777836; PubMed=12897056;
RA Ariyoshi M., Schwabe J.W.R.;
RT "A conserved structural motif reveals the essential transcriptional
RT repression function of Spn proteins and their role in developmental
RT signaling";
RL Genes Dev. 17:1909-1920(2003).
[9]
CC -!- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBP5UH,
CC which prevents the association between NOTCH1 and RBP5UH, and
CC therefore suppresses the transactivation activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTAL1. Interacts with
CC RBP5UH; this interaction may prevent the interaction between
CC RBP5UH and NOTCH1. Interacts with the nuclear receptors RAR and
CC PP4. Interacts with RAR in absence of ligand. Bind to the
CC steroid receptor RNA coactivator SRA.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
CC spleen and thymus. Expressed at intermediate level in kidney,
CC liver, mammary gland and skin.
CC -!- INDUCTION: By hormone 17-beta-estradiol (E2).
CC -!- DOMAIN: The RID domain mediates the interaction with nuclear
CC receptors (By similarity).
CC -!- DOMAIN: The SPOC domain, which mediates the interaction with
CC NCOR2, is essential for the repressive activity.
CC -!- SIMILARITY: Belongs to the Spn family.
CC -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -!- SIMILARITY: Contains 1 SPOC domain.
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF3556524; AA52750.1; -

FT	DOMAIN	1	574	DNA-BINDING.
FT	DOMAIN	6	81	RNA-BINDING (RRM) 1.
FT	DOMAIN	336	416	RNA-BINDING (RRM) 2.
FT	DOMAIN	439	514	RNA-BINDING (RRM) 3.
FT	DOMAIN	518	590	RNA-BINDING (RRM) 4.
FT	DOMAIN	559	575	COILED COIL (POTENTIAL).
FT	DOMAIN	822	850	COILED COIL (POTENTIAL).
FT	DOMAIN	1185	1206	COILED COIL (POTENTIAL).
FT	DOMAIN	1509	1544	COILED COIL (POTENTIAL).
FT	DOMAIN	1607	1627	COILED COIL (POTENTIAL).
FT	DOMAIN	2216	2704	RID.
FT	DOMAIN	3478	3644	SPOC.
FT	DOMAIN	2138	2462	INTERACTION WITH MSX2.
FT	DOMAIN	2706	2845	INTERACTION WITH RBPSUH.
FT	DOMAIN	125	277	ARG-RICH.
FT	DOMAIN	236	326	TYR-RICH.
FT	DOMAIN	648	721	TYR-RICH.
FT	DOMAIN	702	832	ARG-RICH.
FT	DOMAIN	2101	2233	ALA-RICH.
FT	DOMAIN	2377	2518	PRO-RICH.
FT	DOMAIN	2950	3475	PRO-RICH.
FT	VARSPLOC	618	640	Missing (in isoform 2).
FT				/FTid=VSP_008564.
FT	VARIANT	348	348	I -> T.
FT	VARIANT	762	762	S -> F.
FT	VARIANT	773	773	S -> F.
FT	VARIANT	933	933	S -> L.
FT	CONFLICT	754	754	R -> G (IN REF. 3).
FT	CONFLICT	1524	1524	D -> A (IN REF. 3).
FT	CONFLICT	1560	1560	H -> Y (IN REF. 3).
FT	CONFLICT	1570	1570	F -> L (IN REF. 3).
FT	CONFLICT	1574	1574	R -> G (IN REF. 3).
FT	CONFLICT	1609	1609	Q -> R (IN REF. 3).
FT	CONFLICT	1659	1659	I -> V (IN REF. 3).
FT	CONFLICT	1669	1669	S -> F (IN REF. 3).
FT	CONFLICT	1705	1705	V -> A (IN REF. 3).
FT	CONFLICT	1815	1815	A -> V (IN REF. 3).
FT	CONFLICT	2097	2097	G -> A (IN REF. 3).
FT	CONFLICT	2201	2202	MISSING (IN REF. 3).
FT	CONFLICT	2322	2322	A -> V (IN REF. 3).
FT	CONFLICT	2385	2385	P -> Q (IN REF. 3).
FT	CONFLICT	2502	2502	R -> K (IN REF. 3).
FT	CONFLICT	2505	2505	E -> K (IN REF. 3).
FT	CONFLICT	2519	2519	D -> N (IN REF. 3).
FT	CONFLICT	2554	2554	T -> S (IN REF. 3).
FT	CONFLICT	2679	2679	LVSTAGPWN -> VGEHPWARD (IN REF. 3).
FT	CONFLICT	3010	3010	L -> P (IN REF. 3 AND 5).
FT	SEQUENCE	3644	AA; 398750 MW; 9C7EC49A81A7DA4A CRC64;	
Query Match			3.0%; Score 434.5; DB 1; Length 3644;	
Best Local Similarity			18.7%; Pred. No. 1.2e-08;	
Matches 524; Conservative 321; Mismatches 894; Indels 1069; Gaps 128;				
QY	727	RNLNMRPAK-YKDANIMSPGSSLP	SLHVRKQKALELDAQHLSETFDNIDNLSPK--A 783	
Db	226	REVRGRPRPSVQHSRSRSPHSS	-----SQRNSQSPQLA 259	
QY	784	SHRSQRKHQSLGNYGVDFDTNRHDDNRDNFNNTGNMTVLS	PVLTNTVLPSSSSSRGSLDS 843	
Db	260	SQASRPTSPSGSG	-----SRSRSSSDSISSS-----SSSNTDSSDS 298	
QY	844	SRSEKDRSLERBGICLGNYPATNPOTS	KRGILQISTTAAQIAKMWEE-----893	
Db	299	SSTASDDSPAR	-----SVQSAVPAPTSQLLSLDEKDEPRKSFGIK 339	
QY	894	-----VSAIHTSQED	-----RSSGSTT--SLHCVTDERNAL-----RRSSAAHTGNT 934	
Db	340	VQNLVPRSIDTSLKDGLPHEFKY	GKVTISVHIGASERYGLVFRQODEKALTASKG 399	
QY	935	YNF	-----TKSN-----SNRTCSM-----949	
Db	400	KLFFGQIEVTAWVGPETESENEFRP	DERIDEFHPKATRTLFICNLEKTTYHDLRNIF 459	

QY 950 -----PYAKLEY-----KSSNDLSNSVSDGYSKRGQWPK 981
Db 460 QRFGEIVDIDIKVNGVPOYAFLOYCDIASVCKAIAKWDGYLGNRLKLGFGK---SMP 516
QY 982 S-----IESYSDDDESK-----FCYGOY-----PADLAHIIHSAN 1012
Db 517 TNCVMDGLSLSNVSDQYLTHFRFYGVVVFVORLGMALVYSEIEDAQAAYKTKGR 576
QY 1013 HMDNDGELD-----TPINYSKYSDQL-----NSGROSPSONERW---ARP 1052
Db 577 KIGNKIKVDPANRESQAFTHCKEKSQDWDYFEMTERRAQMAQSKHEDWSADAQS 636
QY 1053 KHIIDEIKQS-EORQSR-----NOST--TYPVYTESTDDKHLKQPHFGQOECVPSYRSR 1105
Db 637 PHKCREERRGYSOERTYENVRTPCTYF-----EDSRDYPAR 677
QY 1106 GANGSETNRVSGNHCINQVQSLOCEDDYED---DKPTNYSE-----RYSEE 1150
Db 678 G-----REFYSEWETYQGEYDSRYDEPREYREYRSDPYEQDIREYSYR 722
QY 1151 ECHEEERPTNYSIKYNEKEKHVDOPIDYSLK-----YATDIPS-SOKQS 1194
Db 723 QREERERE-----RFSDRCHERPERSQSPVHLRERPOGVSYPASHERLPDSERRL 777
QY 1195 PSFSKSSSGQSSK-----TEHMSSENSTP-----SS 1223
Db 778 YRSSERSGSCSVSPRYDKLEKALERTYTKNEKADKERTFDPVERERRRIVRKEKGE 837
QY 1224 NAKRONQLHPSAQSRSQPOKAATCKVSSINOETIQYCVEDTPICFSRCSLSLSA 1283
Db 838 KDKAERQKRGKAHSPSQP-----SETEQENDREQSP-----EXPRG 875
QY 1284 EDEICNOTTOADSANTLQ-----IABIKGICTRSAEDPVSEVPVAVSHQHTKSSRL 1337
Db 876 STKLSRDRADKEGPAKRLVPCVVLTRVKEKGVIEHPPEKRLKARLGRDITKASAL 935
QY 1338 -----CGSSLSSESARHAYE---FPSGAKSPSKSG-----AQTP----- 1369
Db 936 DOKPOAAQEPKAPKDPARKALREKVLFSHAEVGEKERTKURKULKAQTPELSALDLE 995
QY 1370 -----KSPPEHYVOETPLMFSR----- 1386
Db 996 KLEARKRPADSLKIEKQPEIKTSPB---TEDTRILLKQPDTRSDGVLLREGESERK 1053
QY 1387 -----CTSVSLDSFESRIASSVQSEPCSMVSGIISPDLPSP 1427
Db 1054 PYRKEILKRESKTKLERLNSALSPKQDPAVASGSGRPSFSDVHAGL---GELTHGS 1110
QY 1428 GQTMPPSRSKTPPPPTQATQKREVPKXKAPAEKRESGPKQAAVNAAVQVQLPDAIT 1487
Db 1111 VETQETOPKAIKPSKQPKQL--OLLENQGP--EKKE-----VRKNYCRPREP 1155
QY 1488 LHFATESTPDGFCSSLSLSALSLDEPFTQKDVLRIMPPVOENDNGNETSEQPKESNE 1547
Db 1156 ABRAGQEKPHGNAEKLGI-----DID-----HTQSYRKQEQSRKQRM 1198
QY 1548 NOEKEAETIDSEKDLDDSDDDIE---ILFECIISAMPTKSRKGGKPAQKASKLPPP 1604
Db 1199 METAKAEFGSGPKOV-----DDYERRSLVHEV-----GKPPQDVTDDSPS 1240
QY 1605 VARKPSQPVYKLLPQNR-----LQPO-KHVSFTPGD---MPRVYCV 1644
Db 1241 KXERTDHDV-FDICTKRENYRSSROI SEDSERTSCSPSVRHGSPHDDDDPRGSPRLVSV 1299
QY 1645 EGTP-----INFSTATSLDLTIESP--- 1665
Db 1300 KGSPPKDEKGLFYPNNAVDRDPLKCNPYDSGKREQTADTAKIKLSVLNSEGEPSRWDDPM 1359
QY 1666 -----FNELAAGEVGRGAQSGEFKRDITPTBGRS-TDE-----AQGGKTSR- 1707
Db 1360 KQDPSRDFSVFNFVNIKRLSLR-----KRSVRDLEFGVPSDDEDAHRQSFRASSF 1413
QY 1708 VTIPELD-----DNKAERGDILAEINCAMPKGSKHKEPRVKIMDQ-----VQOASAS 1756

Db 1414 YDSPRLSFLLRDRDQKLREDD---ERLASSLERKNFYS-FALDKTITPTTKALLERAKSL 1469
QY 1757 SSAPNKN-----QLDGKKKKTSPVKPIPONTEYRTRVRKNADSK----- 1796
Db 1470 SSSREENWFLDWSDFANFRNNKDEKVDSPAPRIPSPWYMKKKKIRTOSEGLDKDKDB 1529
QY 1797 -----NNLNAERVFSDN---XDSKK-QNLKNNSKDFNDKLPNNEDVRGSPAFDPSH 1844
Db 1530 RREEQERQELFASRFLHSSIPEQDSKRIQLHERKSEE-SDPPPG---RLYGRQASEGAN 1585
QY 1845 HYTPISGTYCF--SRNDSLSLDFDDDDVDLSREKAELEK---AKENKESE 1891
Db 1586 STSDSVQEPVVLPHFSFMELTRMQQKEKEKQKPEAEKQESPEHPKTPPEPAETKEPE 1645
QY 1892 AK-----VTSHTBELTNSQOAA-NKTQAAKQPINRG---QPKPILO----- 1928
Db 1646 PKAPVAGLPAVTITVTPPEPASSAPEKAEAEAAEASPAGEKPAEPAPVSEETKLVSFV 1705
QY 1929 -----KOSTFP--OSSKDIPDRGAATDEKIQNFAIENTPVCFNSH----- 1966
Db 1706 SVPVEQPROSDVPPGDSRDSQDAAAPSAPOESAATDAVPCVNAEPIITPGTIVSQVES 1765
QY 1967 -----SSLSLSLSDIDQENNNKENEPKETEPPDS--QGEPSKPOASGVA-----PKSFH 2013
Db 1766 SVDPKTSSPOPLSKLTQREAEAEKV---EKDPTTSTEPDATONAGVASEAQPPASED 1822
QY 2014 VEDTP--VCFSRNSLSLSDIDSEDDLQECISSAM--PKKKPSRLKGDNK-----HSP 2065
Db 1823 VEANPPVAAKDRKTNKSKSKTS---VQAAAASVVEKPVTRKSERI--DREKLKSSSP 1876
QY 2066 RNWGGILGE-----DITLDLKDIOQP-----DSEH----- 2090
Db 1877 R-----GEAKULELMEAEKILTRTASKSGGDTEHPSPSLPLSRRRNRVSVATMT 1930
QY 2091 ---GLSPDSNFOWKAI----- 2104
Db 1931 DHESRPAKEPVQPRVTRKRLERLEQEAUVPTTPRGRPPKTRRAEDGHEHKEPA 1990
QY 2105 -----QEGANSIVSSLHOAAAAACLSE-----QASSDSLSLSKSG--ISLGSPPHL 2150
Db 1991 ETPRPAEGWRSRPSQKSAAAAQPGKRGNEQKVEAAEAGAAQASTREGNPKRGEEAA 2050
QY 2151 TPQEEK--PFTSNKGP-----RIL-KPQEKSTLETKIESESKGIGKGVYKSLITG 2201
Db 2051 SEPKDRDRDSTDKSGPDTFVLEVRKPEK--TYKSKRGARSTRSGMDRAAHQORSLM 2109
QY 2202 KVR-----SNSEISGQMKQPLQANMPSISRGRWMIHIGVRN-----SSST 2243
Db 2110 AARAAGQAADKEAGPAAASPOESESPOKSGSS---POLANNPADPDREAEESESAST 2165
QY 2244 SP-----VSKKGPPKTPASKSPSEGTATTSPRGAKPSVKSELSPV 2285
Db 2166 APPEGTOLARQIELEQAVQNIAKLPEPSAAAAKGTG---TATAASEEPAPHEGHKP- 2221
QY 2286 ARQTSQ-----IGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPG----- 2325
Db 2222 AHQASETELAAGIISDASGEPPENFAPSVP--PGSQTHPREGMEPGLHEAESGILE 2279
QY 2326 ---RNSISGRNGISPNKLSQLPRTSSPSTASTKSSGSGKMSYTSPPGRMSQOINTKO 2381
Db 2280 TGTATESAPQVSALDPPEGSADTKETRGNSGVQVQAKGKAEVTPPRDKRGKQKTRR 2339
QY 2382 TGLSKNASSIPRSEBSASKGLNQMNNGANKVKELSRMSSTKSSGSESRSRPERVLRQS 2441
Db 2340 -----RKNANKV--VAITETRASEAEQTOSESPAASEAT 2373
QY 2442 TFIKEASPTLRRKLESASFESLSPSRPASPTRSOAQTPVLSPLPDMSLTHSSV-- 2499
Db 2374 AATPEAQ-----BE-----KPSEKPPSPAECTFPDKTP--PSSLSOENSAE 2417
QY 2500 ---QAGWRKLPNLSPTEYNDGRPA--KRHDIAHSHESP-----S 2537

Db 551 DTGLG-----DKPFLD-----TAEAGPPSTAIQOTPPSVPLGQEE 587
Qy 956 YKRSNDSLSVSSNDGCKGQKMPKSIESYSEDDSKFCSYQYPADLAHITHSANMD 1015
Db 588 HVMKEKELVVPVEEQSGKRG-----LDSGAETEEK----- 620
Qy 1016 DNDGELDTPINYSIKYISDEQLNSGRQSPSQNRWARPKHIITEDBIKOS--EQORSNQST 1073
Db 621 -----DI-----WEEKQREARLPDRTE--AREES--EPEVKEDVIEKAELEEMEE 663
Qy 1074 TYPVYTESTDDKHUK--POPHGQOECVSPVRSRGANG-----SETNRVCN- 1118
Db 664 VHPDEEBEDATKABGFQKIMQPLKVTP--RSREAFGREGLGQKAPKETSLFLSSL 722
Qy 1119 --HGINQVNSQSLCQEDDYDDKPTNYSE--RYSEBEOHEE--EERP-----TNYSI 1164
Db 723 TTPAGATEHVSY-----IQDEIIPGSETEQIISDEIHDEEERAPPRFHTSIYDL 775
Qy 1165 KYNBEKR--HVDQPIDYS-----LKATDI--PSSOKQSPFSKSSSQG 1204
Db 776 PQEGAGFEASQPADSAPVATSGKVGTPETELTYPTNIVAAPLABEEHVSSATSITEC 835
Qy 1205 SSKTEHMSSENSTPSSNAKQNLHPSSAQSRSGOPOKAATCKVSSINQETIQYCV 1264
Db 836 DKLSPFATSVABEDQSVSLTA-----PQTEETGK--SSLLDITV----- 872
Qy 1265 EDTPICFSR-----CSSLSLSAEDBIG-----CNQTTQPADSANTLQIAEI 1307
Db 873 --TSIPSKRTBATQGLDVVPSAGTIPSTPSLEEDKGFSPCEDFVTVGESEKRGELIIG- 929
Qy 1308 KKGITGRSAED-----PVSEVPVAVSOHPRTKSSRLOGSSLSSESA 1347
Db 930 KGLSGERAVEBEEETANVEMSEKLCQYGTGTFVPSAPGHALHP-----GEPA 976
Qy 1348 RHKAVE--PPSGAKSPSKGAOTPKSPRHHYQVETPLMFESRCTSVSSLDSFESRSIASSVQ 1406
Db 977 LGEABERCJLSPDDSVKVASPPSPGPPS--ATHIPPHQSPVEEKSEBPQDFQAD----- 1028
Qy 1407 SEPCCGMVGIIISPDLPDSFGQTNWPPRSKTPPPPPQ--TAQTKREVPPKNAKTAEKRE 1464
Db 1029 -----SMGDTKRTFGVGKEDAAEETVKPGPEEGTLEKEEKVPPPRSPQAQEP 1076
Qy 1465 SGPKQAAVNAVQVQLPDADTLHF-----ATESTPDGFSC----- 1502
Db 1077 VNIDBGLTCT--IQLPDAQDKALVFEIMEAGEPTGILGABALPGGLRLPQEPGKQ 1133
Qy 1503 -----SSLSALS-----LDPEFTIQDVELRIMPPV 1528
Db 1134 KDEVLRYPDRLSPEDAESLSVLSVPSPTANOEPKSPCGLTEQYLHKD-----RWPEV 1189
Qy 1529 QENDNGNETESQPKESNENQEKAEKTIIDSEK-----DILDD 1566
Db 1190 SPEDTQSLSGS-----ESPSETSLDVSSKQUSPESLGTQLQFELNLGKEEMGHLMQA 1243
Qy 1567 SDDD-----DIEILERICIASMP-----TKSSRRKGGKPAQTASKLPPPVARK 1608
Db 1244 EDTSHHTAPMSVPEPHAATASPTDGTTRYSAQTDITDSDLRKSPPASSFSHTP--SGN 1301
Qy 1609 PSQLPVYKLLPSQNLQOKHVSFTPGDDMPRVYCVGTEPINF--STATSLDILIE----- 1663
Db 1302 GRYLFGAITSDEHILTPDSSFSKSP--ESLFGP--ALEDIAIKWEDKVPGLKDRTSEQKE 1359
Qy 1664 -SPPNEL-----AAGEVGRGAQSG--EFEKRD 1698
Db 1360 PFPEKDEVLOQDKTLEHKEVVEPKDTALYQKDEALHVKNVAVKQDKALEQKGRDLEQD 1419
Qy 1689 T-----IPTGEGSTDBAOGGKTSVITIPELDNKAEBEGDILAEACINSAM--PKGSKHX 1739
Db 1420 TALEQKQKALEPKDKLEBKQKALEQKQKIPBEKDKALEQKQKALEQKQKALEPKDKDLE 1479
Qy 1740 -----PFRVKIMDQVQAASASSASAPKNQOLDKKKKPTSPVKIPONTVEYRT 1787
Db 1480 QKDRVLEQKEKIPBEKQKALDQ--KVRSEVHKAPEDTVVEMKORDLEQKTDKAPBQHQAOE 1538

Qy 1788 RYRKNAADSNNLNAERFVS--DNKDSKKON--LKNSKDFNDKLPNNEDRVRGSPAFD- 1841
Db 1539 QDKVQSEKKQDALEQKYWALQKQDEALEQNIQALENHQTOQEESLVQEDTKPKWLEB 1598
Qy 1842 -SPHYTPIEGTPYCFSRNDSLSLDDDDVDLSREKAELRAKENKE- 1889
Db 1599 KSEPKVAME-----EKLEALLEKTKALGLEBSLVQEGRARQEEKYWRGDWQOE 1649
Qy 1890 -SEAKYTSHTELTSNOOSANKTOATAKQPIN--RGQPKPILOKQSTPQSS----- 1937
Db 1650 WQSTSTREBPAGEQKELAPAWEDTSPQDNRYWRGREDVALEQDITYWELSCERKWWFP 1709
Qy 1938 KQIDPRGAA--TDEKLQNFAIENTPVCFSHNSLSLSDIDQENNNKENETIKE--TBP 1992
Db 1710 HELDGGQARPHYTEERESTF-----LDEGPDDEQEVPLREHATRS 1749
Qy 1993 PDSQG-----EPSPQASGAPKSPHYE-----DTPVCFSRNSLSLSISDEDDLQ--E 2041
Db 1750 PWASDFKDFQESSPO-----KGLEVERWLASPV-----GLPPEEDKLTSPPE 1794
Qy 2042 CIS-----SAMPKKKPKSRLKGDNEXHSRPNMGILGEDTLTD-LKDI-----QRPDSE 2089
Db 1795 IISPPASPEMVQORVPSAPGQESPIPDPKLPHMKNEPTTTSWLADI PPWVPKDRPLPP 1854
Qy 2090 HGLS-----PDSN--FDW-----KAIQEGA-----NSIVS 2113
Db 1855 APLSPAPGPPTAPAFESHTTAPFSGWGTPEYDSVVAQVQGAALLEGPGYPLGSKDYKXVG 1914
Qy 2114 SLHQAAAAACLSQAASSDSDSILSLKSGISLGSPPHLTPDOER-KPFTENKG-----P 2165
Db 1915 EREBERGAAPDK--SSHXKVPPEARKSHATTEPEQTEPEQREPTYPDERFOYADIYE 1972
Qy 2166 RILKPGKSTLETKKTIESKGIKGGKYKYSIITOKVRSNBSISQMKQP-----LOAN 2220
Db 1973 QMWLTGLGPACPTRE--PPLGAAGDWPPCLSTKEAAAGRNTSAEKELSPISPKSLQSD 2029
Qy 2221 MPSIS-----RGRM--IHIPGVNSSSTSPVSKGPPKLPASKSPSGQATTTSPRGA 2274
Db 2030 TPTFSVAALAGTVPVPRXEPGSPMBESLTPFVAPPAPILISKGPSPLNGNLIUSCPDRR 2089
Qy 2275 KPSVK-----SELSPVARQTSIQIGSSKAPSRSGRSDSTPSRPAQOPLSRP- 2320
Db 2090 SPSPKESGRSHWDDSTSDSELEKGAQEAE-----KEAQSPSPHPPIPGSPT 2137
Qy 2321 -----TQSPGRNISPGRNOISPP-----NKLSQLPRTSSPSTASTKSGSGKMSY 2366
Db 2138 LWPETEAHVSPPLXSHLGKARPSLDPFASAFGFSLEXAP-POXPSPAPRSPAPCGSLAF 2196
Qy 2367 T-----SPG-----RQMSQQLTKOTGLSKNASSIPRSES-ASKGLNOMNNGANGANK 2413
Db 2197 SGDRALALAPGPTRTYDEYLEVTKAPSLD---SSLPOLPSFSSPGXPPLSN----- 2246
Qy 2414 VELSRMSTKSGSGESDRSERPVLVRSQSTFIKAPSTTLRKLUESASFE----- 2463
Db 2247 --LPRPASPALSGSGSSEATTPI---8SVAERFSPSLEAARQESGELDPGMEPAHXL 2300
Qy 2464 ---SLSPSRPASPITSQAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIET-- 2517
Db 2301 WDLTPLUSPAP-PASLDLALAPASLPGDMDGDLPLCHLECESEAAATEKPSFPQVPSDECAA 2359
Qy 2518 -----NDGRPAKRDHIARSHSPSRPLPNRSQTW-----KREHSKHSSSLPRYSTWR 2565
Db 2360 NGPTETSPNPPXPAP---AKAENEZAAAXFAWERGAWPEGAERSRPRDXTLSPEQVPCP 2415
Qy 2566 RTGSSSILLSASSE---SSEKAKSEBKHVNSISGT-----KQSKENQVSA 2608
Db 2416 XGSGGGPPSSASPEVEAGPQGCXTEPRPHRKGELSFNLPPLPPSIDDRLDLSTEEVRLVG 2475
Qy 2609 KGTWRKI--KENEFSPNTNSTQTVSSGATNGABSKTIIYQMAPAVSKTEDVWVRIEDCP- 2665
Db 2476 RGGERRVGGGTTCXPVPTDETPTPTISASDSGSSQS-----DSDVPPTEECPS 2524

QY 2666 -----INNPRSGRSTGTPPVVIDSVSKANPNKIDSKD 2699
 Db 2525 ITAEALDSDEGDFLPVXXGSGVTHPRPGHDP-----PPL-----PQDXRPS-- 2571
 QY 2700 NQAKQNVGSGVMTVGLNRLTSTFIQVADPQKGTETKPGQNNPVPVSETNESPIVER 2759
 Db 2572 -----PPRP-----DVCMAPEGLSGESGEXRLRXKXKQVGRVGR 2608
 QY 2760 TPRSSSSKHSFSGTVAARVTPFNYPNPSRKSSADSTARSQIPTPWNNTKKRDSK 2819
 Db 2609 AP-----CKDPVFXRLXLR-----GKSFPGSGXDRVSRXPSRSTQSVTPAEEXD 2661
 QY 2820 TDSTESSG 2827
 Db 2662 GHSPMSXG 2669

RESULT 10
 PCLO_RAT STANDARD; PRT; 5085 AA.
 AC Q9UKS6; Q9JLT1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
 GN PCLO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
 RX MEDLINE=20170257; PubMed=10707984;
 RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
 RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
 RT "Piccolo, a presynaptic zinc finger protein structurally related to
 RT bassoon.";
 RL Neuron 25:203-214 (2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694
 RX MEDLINE=21181819; PubMed=11285225;
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RA "An unusual C(2)-domain in the active-zone protein piccolo:
 RT implications for Ca(2+) regulation of neurotransmitter release.";
 RL EMBO J. 20:1605-1619 (2001).
 CC -I- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -I- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
 CC similarity).
 CC -I- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC IsoId=Q9UKS6-1; Sequence=Displayed;
 CC IsoId=Q9UKS6-2; Sequence=VSP_003930, VSP_003931;
 CC -I- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -I- SIMILARITY: Contains 1 PDZ domain.
 CC -I- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF138789; AAF07822.2; -;
 DR EMBL; AF227534; AAF63196.1; -;
 DR HSSP; P04410; 1A25.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
 DR GO; GO:0005522; F:profilin binding; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR01478; PDZ.
 DR InterPro; IPR008899; Znf_piccolo.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF05715; Zfpiccolo; 2.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.
 DR PROSITE; PS0106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT P-A-K-P-Q-P-Q-P-X. 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRRK (in isoform 2).
 FT FTID=VSP_003930.
 FT Missing (in isoform 2).
 FT /FTID=VSP_003931.
 FT D-A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT D-A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT V-S: SMALL INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT M->S: INCREASED AFFINITY FOR CALCIUM.
 FT VW->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT QN->AA: MODERATE INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT A-S: NO EFFECT ON CALCIUM-BINDING
 FT ACTIVITY.
 SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
 Query Match 2.7%; Score 400; DB 1; Length 5085;
 Best Local Similarity 19.9%; Pred. No. 3.3e-07;
 Matches 387; Conservative 243; Mismatches 765; Indels 546; Gaps 83;
 QY 1187 IPSSQKQSPFSKSSSQSSKTEHMSSTSTSTPSNAKRNQQLHPS--SAGSRSGPQ 1244
 Db 67 VPPAAASPSMHRKQELDSQAQPGKPPDPGRPTQPGLSKSRRTDTFRSEKLPGRSP 126
 QY 1245 KAATCKYSSINQETIQYCYVEDTPIFCRCSLSLSLSAE-----DETGCNQTQADS 1298
 Db 127 STSLKSKSGRTDFKESYKSMMPGFFSDVNPUSAVSSVNVKNPFDLISDSASQETT 186
 QY 1299 ANTLQIAEIKGKTGTRSAEDPV---SEVPASVQHPRTKSSRLQGS---SLSSSARHKAV 1352
 Db 187 KKQKVQKEQK-SEGMAKPPLOQSPKPP-KPQGGQVKEVIOQDSPPKSVSSQAQKVP 245

Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

(3)

SEQUENCE FROM N.A. (ISOFORM 1).

TISSUE=Brain stem; PubMed=8253844;

RA Chan W., Kordeli E., Bennett V.;

RA "40-kD ankyrinB: structure of the major developmentally regulated

RT domain and selective localization in unmyelinated axons.";

RL J. Cell Biol. 123:1463-1473(1993).

[4]

SEQUENCE OF 463-495 FROM N.A.

RA MEDLINE=92009921; PubMed=1833308;

RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,

RA Lux S.E., Ward D.C., Forget B.G.;

RT "Isolation and chromosomal localization of a novel nonerythroid

RT ankyrin gene.";

RL Genomics 10:858-866(1991).

CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal

CC elements. Also bind to cytoskeletal proteins.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Name=1;

CC IsoId=Q01484-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;

CC Name=3;

CC IsoId=Q01484-3; Sequence=VSP_000268;

CC -!- TISSUE SPECIFICITY: Plasma membrane of neurons as well as glial

CC cells throughout the brain.

CC -!- PM: Phosphorylated at multiple sites by different protein kinases

CC and each phosphorylation event regulates the protein's structure

CC and function (Potential).

CC -!- SIMILARITY: Contains 23 ANK repeats.

CC -!- SIMILARITY: Contains 1 death domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL; X56957; CAA40278.1; -

DR EMBL; X56958; CAA40279.2; -

DR EMBL; Z26634; CAB42648.1; -

DR EMBL; M37123; AAA62828.1; -

DR PIR; S37431; S37431.

DR HSP; P42771; IDC2.

DR Genew; HGNC:493; ANK2.

DR MIM; 106410; -

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000488; Death.

DR InterPro; IPR000906; ZU5.

DR Pfam; PF00023; ank; 24.

DR Pfam; PF00791; death; 1.

DR PRINTS; PR01415; ANKYRIN.

DR SMART; SM00248; ANK; 22.

DR SMART; SM00005; DEATH; 1.

DR SMART; SM00218; ZU5; 1.

DR PROSITE; PS50088; ANK REPEAT; 20.

DR PROSITE; PS50297; ANK REP REGION; 1.

DR PROSITE; PS50017; DEATH DOMAIN; 1.

DR Cytoskeleton; Alternative splicing; Repeat; ANK repeat;

KW Phosphorylation.

FT REPEAT 63 92 ANK 1.

FT REPEAT 96 125 ANK 2.

FT REPEAT 129 159 ANK 3.

FT REPEAT 162 191 ANK 4.

FT REPEAT 193 220 ANK 5.

FT REPEAT 232 261 ANK 6.

FT REPEAT 265 294 ANK 7.

298 327 ANK 8.

FT REPEAT 331 360 ANK 9.

FT REPEAT 364 393 ANK 10.

FT REPEAT 397 426 ANK 11.

FT REPEAT 430 459 ANK 12.

FT REPEAT 463 492 ANK 13.

FT REPEAT 496 525 ANK 14.

FT REPEAT 529 558 ANK 15.

FT REPEAT 562 591 ANK 16.

FT REPEAT 595 624 ANK 17.

FT REPEAT 628 657 ANK 18.

FT REPEAT 661 690 ANK 19.

FT REPEAT 694 723 ANK 20.

FT REPEAT 727 756 ANK 21.

FT REPEAT 760 789 ANK 22.

FT REPEAT 793 822 ANK 23.

FT DOMAIN 1773 1950 REPEAT-RICH REGION.

FT REPEAT 1773 1784 REPEAT A.

FT REPEAT 1785 1796 REPEAT A.

FT REPEAT 1797 1808 REPEAT A.

FT REPEAT 1809 1820 REPEAT A.

FT REPEAT 1821 1832 REPEAT A.

FT REPEAT 1833 1844 REPEAT A.

FT REPEAT 1845 1856 REPEAT A.

FT REPEAT 1857 1867 REPEAT A.

FT REPEAT 1868 1879 REPEAT A.

FT REPEAT 1880 1891 REPEAT A.

FT REPEAT 1892 1902 REPEAT A.

FT REPEAT 1903 1914 REPEAT A.

FT REPEAT 1915 1926 REPEAT A.

FT REPEAT 1927 1938 REPEAT A.

FT REPEAT 1939 1950 REPEAT A.

FT DOMAIN 3536 3620 DEATH.

FT VARSPLIC 1039 1039 Q -> QFLGKHLPTAPPPLNEGESLVSRILQLGPPGTG (in isoform 2).

FT VARSPLIC 1444 3528 /FTID=VSP_000267.

FT MISSING (in isoform 2 and isoform 3).

FT CONFLICT 475 476 Q -> PE (IN REF. 4).

FT CONFLICT 971 971 I -> S (IN REF. 1).

FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).

FT CONFLICT 3586 3586 I -> Y (IN REF. 1).

SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;

Query Match 2.7%; Score 395; DB 1; Length 3924;

Best Local Similarity 19.5%; Pred.No.3.7e-07;

Matches 427; Conservative 307; Mismatches 799; Indels 658; Gaps 111;

QY 897 IHTSQEDRSSGTTTLHCVTDERNALRRSSAATHGNTYNTKSE--NSNRTCSMPYAKL 954

DB 1421 IYTKESSEDQEQBEID-WTSEKN-----DETETSVLKSHLVNEVPVLASDLS 1472

QY 955 EYKSSND--SLNSVSSNDGYGKRG--QMKPSTESYSEDDSEPCSYGQVPADLAHKHS 1010

DB 1473 EVSEMKQDLIKMTAILTTDVSXKAGSIKVELKAAEEEEEPGF----- 1516

QY 1011 ANHMDNDGELDTPIYSLKYSDEQLNSG-----RQSPQ---NERWARPKHIE 1057

DB 1517 -----EIVERKEDLEKVEILASGTCTDESSVQSSRSRGLVVEEWV-----IVS 1563

QY 1058 DETKQSEQRNRQSTTPYVYTESTDDKHLKFOHFGQECVSPYRSGANGETNRVGS 1117

DB 1564 DEETEEARQKAPLEITEYPC-VEVRIDKTEIK-----GKVE---KDSGLVNYLTDDLNT 1613

QY 1118 NHGINQNVSSQLCEDDYDDKFTNYSE-----RYSEEQHEEEERTPTNYSIKYNEKSH 1172

DB 1614 CVPLPKEQLQTV-----QDKAGKCEALAVRSSEKKGKIPPDDETQSTQHKPISLG 1666

QY 1173 VDQPIDYSLKYATIDIPSSQKQSFPSKSSSGQSKTEHMSSESSTSTPSSNAKRONQLH 1232

DB 1667 IKKPVRRKLK-----EKQKQ-----KEEGLQASAKA-----ELK 1696

QY 1233 PSSAQSRSGQ-----PQKATCKVSSINQETIOTYCVETPTICFSRCSLSLSLSAED 1287

J. Cell Biol. 147:151-162 (1999).
 [2]
 RN SEQUENCE OF 552-4404 FROM N.A.
 RA Kraemer J., Wollam C., Wohldmann P., McGrane B.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
 RP TISSUE=Brain;
 RC MEDLINE=98290545; PubMed=9628581;
 RX Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:31-39 (1998).
 [4]
 RN SEQUENCE OF 4405-4439 FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=22388257; PubMed=12477932;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dapkin L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
 RA Richards S., Worley K.C., Harte S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [5]
 RN SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Y19188; CAB60727.1; -.
 DR EMBL; AC004903; AAD20936.1; -.
 DR EMBL; AC004886; AAD21789.1; -.
 DR EMBL; AB011131; BAA25485.1; -.
 DR EMBL; BC001304; AAH01304.1; -.
 DR EMBL; AC004082; AAB97937.1; -.
 DR PIR; T06634; T06634.
 DR HSSP; P04410; 1A25.
 DR MIM; 604918; -.
 DR GO; GO:0005856; C:cytoskeleton; NAS.
 DR GO; GO:0045202; C:synaptic junction; ISS.
 DR GO; GO:0005509; F:calcium ion binding; ISS.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
 DR GO; GO:0005522; F:profilin binding; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 DR InterPro; IPR000008; C2.
 DR PRINTS; PR00360; C2DOMAIN.
 DR SMART; SM00239; C2; 2.
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT NON_TER 1
 FT DOMAIN 400 465
 FT ZN_FING 499 523
 FT ZN_FING 969 992
 FT NON_CONS 1010 1011
 FT DOMAIN 2300 2325
 FT DOMAIN 4391 4442
 FT DOMAIN 4544 4633
 FT DOMAIN 5031 5121
 FT VARSPLIC 4404 4404
 FT VARSPLIC 4534 4534
 FT VARSPLIC 4576 4576
 FT VARSPLIC 4757 4761
 FT VARSPLIC 4762 5147
 FT SEQUENCE 5147 AA; 563537 MW; C5D587990496CD3C CRC64;
 Query Match 2.7%; Score 387.5; DB 1; Length 5147;
 Best Local Similarity 19.3%; Pred. No. 9.6e-07;
 Matches 506; Conservative 313; Mismatches 961; Indels 837; Gaps 120;
 QY 718 IAGSAAALNLMANRPAPKYNKDNIMSPGSSL-----PSLHVKKKALEALD 765
 DB 1 IPAGNEADLSQ--SEERQIAAVMSRAOGLPKGVFPAAASPMSH-RKQ-----ELD 52
 QY 766 AQHLSSETFDNLSLSPKASHRSKQRIKQSLYGVDFDTRNHRDNRSDNFMNTGLSPY 835
 DB 53 SSH-----PPKQSGRPP-----DPGRPAQGLSKSRITDTF 83
 QY 826 LNTVLPSSSSSRGSLDSSSEKDRSLRERGLGNYPATENPG-TSSKRGLOISTTA 884
 DB 84 RSEQLPGRSPFTISLKSRSRTD-----LKEHKSSMPGFLSEVNALSAVSV 133
 QY 885 AQIAKVMEEYSAIHTSOEDRSSGSTTELHCVTDE----- 918
 DB 134 VKRNPFDLISDSEASQEE-----TTKKQKVQKQKPEGIKPQQPPKPKQOG 188
 QY 919 --RNALRRSAANTHTSNYNTFTKSENRTCSNP-----YAKLYKRS 959
 DB 189 PGRDPLQDQDTPKXIS-----SQQPEKIKSPGTGKPIQGTPTPTDTHAKLPQR- 240

QY 960 SNDSLNSVSSNDGYKGRQMKPSTESYSEDDSKFCYGOVPADLAHAKHSANHMDDNDG 1019
Db 241 --DASRQTKQADIVRGESVXPSLPSKPI-----QPTPKPAPQOPGHEKSPQ 291
QY 1020 ELDTPIYNS-----LKYSDEOLNSGRSPSONERWAPKHIIEDEIKOSBORQSRNOSTT 1074
Db 292 PAKPPAPQSGLTPLAQCPGVKPPVPPGPTTKPPAQP--LGPAPKPPAQQTGSEKPSSE 348
QY 1075 YPVYTESDDKHLKFQPHFGQOECVSPVRSRGANGSETRVGSNHGINQVNSQSLCOEDD 1134
Db 349 QP-----GPKALAQPPGVK-----TPAQQGPAPKPTQOQV-----PKPLAQOPG 390
QY 1135 YEDDK-----PTNYSERVSEEQHEERPTNYISIKYNEEKHVDQPIDYSKYATDIPSS 1190
Db 391 LQSPAKAPGPTKTAQTKPPSQPGSTKPPQ--QPGAPKPSPOQ-----GSTKPPSQ 442
QY 1191 QXQSFPSKSSGGSSKTEHMSSENTSTSSNAKQNLHPSSAQSRSGOPOKATCK 1250
Db 443 QPGS---AKPSAQOPSPAK---PSAQOFTKVPQGTGFGKPLQPPVPSFAKQPP-----490
QY 1251 VSSINQETIQYCVEDTPI-----FSRCSSLSLSSAEDDEICNQTTQEA 1296
Db 491 ---SQGLPKIIC-----PLCNTTELLHVPEKANFNTCTEQTTVCSLGCNPNPHLTEA 542
QY 1297 DSANTLOIABIKGIGTRSAEDPVSEVPAVQCHRTKSSRLQSGSLSSASARHKAFFPS 1356
Db 543 KEWLCLN--COMKRALGGDLAPVPSPOPKLTAPVTTTSAVSKSPQOQTSPPK---KD 597
QY 1357 GAKSPSKGAOTPKSPPEHYQVETPLMFSRCTSVSSLDSPSRSIASSVQSEPCGMVSG 1416
Db 598 AAPQDLSKAPKPKPLVQPT-----LHG 624
QY 1417 IISPSDLPSFGQWPPRSKTPPPPPQATQKREVPKNAKPTAKRESGPKQAAVNAV 1476
Db 625 --SPSAKAP-----PEADSLSKAPPKPEPSVPSE--QDKAPVADDPKQPKM-----669
QY 1477 QRQVLPDADTLHPATESTPDGFCSSSLGALSILDEFFIOKDVLELRIMPPVQENDNGNE 1536
Db 670 -----VKPTDLVSSSATTKPDIPSSKVSQAEKTTTPL-----K 706
QY 1537 TESEQPKES-----NENQEKAEKTIIDSEKDLDDSD 1567
Db 707 TDSAKPSQFPPTGKVTFFDSKAIPFASDASKIISHPGPSESSEKQKV-----756
QY 1568 DDDDEIIEBCLISAMPKSRK-----GKPAQTASKLPP--PVARK-----PSQLPV- 1614
Db 757 --DPVQKKE-----PKAQTKMSKPKDPAKMPKSGSTPPGPRTAGQTVPTPQQSPK 808
QY 1615 -----YKL-----LPSQNRLOPKHVS-----FTPGDDMPRVYC 1643
Db 809 QEQSRPFLNLGSIITDAPKSQPTTQETVTKLFCFGASISFOASNLISITAGQCPHQS 868
QY 1644 VEGTPIFSTATSLDLTIESPNELAAEGVGRGAQSGEPEKRTIITEGRSTDEAOGG 1703
Db 869 GFGAPMKAOPAPS--QPTSQGPBK--STGQAPPAPAKSIPVKETKAPAAEKLEPKAQ 925
QY 1704 KTSVVT-----IPELDDNKAEBGDILACINSAMPK--GKSHKPPRVKIMDOVQOASAS 1756
Db 926 PTVKKTETEKPPINDSKS-----LTAEQKAVLFTKLEKSPKPESTCPLC-KTELNTGS 980
QY 1757 SSAPKN-----QDGGKKKXPTGP-----VKPI 1779
Db 981 KDPNFNTCTECKQVNCNLCGNPTPLTENCOTQRAISGQLGDIRKVPAPSPGPKASPM 1040
QY 1780 PONTYRTR-----VRKNADS-----KNNLNAER-----VFSNKKDSK 1812
Db 1041 PVPTESSQKTAVPQOVKLVKQOEVEKTEAEKILEKVKETLSMEKIPPMVTTTQKQEE 1100
QY 1813 KQNLKNSKDFNDKLP-----NNEDRVGS-----FADSPHHYTPIEG--TPYCF 1857
Db 1101 SKLEKDASALQEKPLPEKKLIPEEKIRSEKKPLLEKKP--TPEDKLLPEAKT 1157
QY 1858 RNDLSLDFDDDDVDLSREKALKA-----KENKESEAK---VTSHTELTSNQOASANKT 1910

Db 1158 SAPEEQKHLKLLSQVQIAEBKLEGRVAPKTVQEGKQPTKMEGLPSGTPOSLEKDDKTT 1217
QY 1911 QAIKQPINRGQPKPILOKQ-----STFPOSSKDIIDRGAA-----1946
Db 1218 KTIKEQFPCTAKPDQEKEDDKSDTSSQQFASQGLSDTGYSSDGISSSILGEIPSLIP 1277
QY 1947 TDE-----KLQNFIAIENTPVCFGHNSLS-----LSDIDQENNNKENEP-----1986
Db 1278 TDEKIDILKGLKDSFQESSPSPDLAKLESTVLISLEAQASTLADEKSEKKTQPHYS 1337
QY 1987 -----IKETEPDSDQGPBK-----PQASGYAPKSPHVED 2016
Db 1338 PEQKQOEKTOSETLEITISEEIKESQEBKOTFKKDSOODIPSSKDKHEKSEFVDD 1397
QY 2017 TPVCFSRNSSLSSIDSDDLLQECISSAMPKKKPSRL-KGDNEKHSFRNMGJLGED 2075
Db 1398 IT---TRFPYDSVERSESE-----NSPVQKERTSVGSSSSDEYKQEDSQSGE 1448
QY 2076 LTLDLKDIOQSEHGLSPDSENFDWKAQ-----GANSIVSLHQAATA 2121
Db 1449 DFIRKQIIEMSADEDASGEDEDEFIRNQLKEISSTESQKKEETKKGKITAKHRR---1505
QY 2122 ACLSROASDSDSILSLKSGISLSPHLPPOEEKPFTSNKGPRIILKPGKSTLTKKI 2181
Db 1506 --JTRKSSISIDEDAGR-----HSHWDEDDFAF--DESP-----ELKYRETQSK 1546
QY 2182 ESEKGIK--GGKVVYKSLITGVRNSI-----SQMKOPL-----QANMPSIS- 2225
Db 1547 ESEELVVTGGGLRREKT-----IELNSTIADKYSAESSQKKTSLYFDEBEPELEMESLTD 1601
QY 2226 -----RGRTMIH---IPGVNRS-----SSTSPVSKGPPKLPKSPASEG-- 2264
Db 1602 SPDRSRGEGSSHLASSPTPGTSPISVSLDSDSSFSHKKGESKQKAKARHPHGL 1661
QY 2265 -QTATTSRPAKPSVKSLSVARQTSIQGSKAPSRSGSRDSTPSRPAQOPLSPIS 2323
Db 1662 LPTIEDSBEELREBEELLKEQKREIEQQKSSKSKKDKDELRAQRRRRPKTP 1721
QY 2324 PGRNSIPGANGSPNKLQ-----LPTS--SPSTASTKSSGSGKMSVTSRQMSQ 2376
Db 1722 P--SNLSPIEDA--SPTEELRQAAMBELHRSSCEYSPIESDEPG--FEISBEKIEVQ 1776
QY 2377 NLTK--QTGLS---KNASSIPRSASKGL-----NOMNGN-----GANKKVEL- 2416
Db 1777 KYVKLPYAVSLYPTDEQSLMOKEGQKALKASAEVMEEMHKTTHYKAPPAANERDEVF 1836
QY 2417 -----SRMSSTKS---SGSESDRSE-----RPLVVRQSTFIKE 2446
Db 1837 EXEPLVGMILIEDYIVESLVEDTYNGSDGLLTOEENGFMQKQREOKILSQQIYE 1896
QY 2447 APS---PTLRKLEESASFESLSPSRPAPSPTRSQATPVLSPS--LPDMSLSTHSSVQA 2501
Db 1897 DPMQKITDLQKREYELSLHVV-----QEDIVSSSFPIPE-----SHEIVDL 1940
QY 2502 GGM-----RKLPPNLSPITIEYNDGRPAKRHDIAKSHSESPSRPLPINRSGTWKRHSKH 2554
Db 1941 GTWVTSTEEERKL--LDADAAYEE--LMKROQMQLTPGSSPTQAPIGEDMT---ESTMD 1992
QY 2555 SSSLPRVSTWRTGSSSSILSASS---ESSEKAKSEDEKHVNSISGTOKSKENQVSAKGT 2611
Db 1993 FDRMPDASL-----TSSVLSGASLTDSTSSATLSIPDVKITQHFSTEEDEYVI--DY 2044
QY 2612 WRKIKE-----NEF-----SPTNSTQTSVSSGATNGAESK-TLYIQMAPA 2650
Db 2045 TREIQEIIAHESLLITYSEPSSESATSVPSPDTFSLTSSVSSVCTTDDSSSPITLDSITT 2104
QY 2651 VKTQEDVWVRIEDCP-----INNPRGRSPRTGNTPP-----2681
Db 2105 YTEPVDMITKFDEEISEISSSTYFPGSIIDYPERISVSLDRTAPDGRASADHIVISUDM 2164
QY 2682 ---VIDSVSEKANPNIKDS-----KD--NOAKQVNGVSPMRTVGLNRLTSFIQ 2727

QY 179 SLOTLDRLEQYEAQIRVAMEBOLGTC-QDMKRAQRIARIQOIEKD-----ILR 230
Db 628 KL---LTQWSEKENLOSKI---NHLETCLKTQOIKSHYENVRVLTLEMDRENLSVEIRN 681
QY 231 IROLLOSOATBAERSS-----QNKHETGSHDAERONQGOVGINMATSGOGSITRM 284
Db 682 LHNVLDSKVEVETOKLAYMELQQAEPFSDOKHOKIEB-----NMCLK-----724
QY 285 DHEATASVSSSTHAPRRLTSHLTKVEMYVYLSMLGTH--DKDMSRTLLAVSSSQD 342
Db 725 -----TSQLTQGVDELHKLQLLSNEIMDKORCYQDLHAYES--762
QY 343 SCISNRQSGCPLLIQLLHNGDKSVLLGNRSGSKEARA-PASAAH-----NIHSQD 397
Db 763 -----LRDLL--KSKDASLTNEDHORSLLAFQOQAMHHSFANIIGEOGSM 807
QY 398 KRGREIRVHLLEQIRAYCETCWEQAEPGMDQKNP-PAPVEHQICPAVCVLMKL 456
Db 808 PSESECR-----LEADQSPKNSAILQNRV-----DSL 835
QY 457 SPDEHRRHAMELGLQIAELLQVDCMYGLTNDHYHITRRVAGMALTNLTGDDVANK 516
Db 836 BPSLESQKMN-----SDLOKQCE-----ELVQIKGEIEENLMKAEQMHQ 875
QY 517 ATLCMKGMCRALVAOLKSESDLOQVIASVLRNLSWRADVNSKKTLEVGSVKALMECA 576
Db 876 SPVAETS-----QRISKLQEDTSAHQNVVAETLSALE-----NKEKELQLINDKVETEOAE 926
QY 577 LEVKEEST--LKSVLSALNLSAHCTENKADICAVDGAFLVGLTYRSOTNTLAIIES 634
Db 927 IQELKXSNHLLSDLSKELQLLS-----ETLSLEK-955
QY 635 GGGILRNYSLSLATNE-DHQIILRNENCLQTLQHLKXSHLTIIVSNACGTLNLSARNPK 693
Db 956 -----KEMSSITSLNKRREIEUTQENGTLKEINASLNQEKWLIQKS-ESPANYIDEREK 1009
QY 694 DOEALWDMGAVMLKNLHSHKMTAMGSAALRNLMANRPAPAKYDANIMSPGSLPLSH 753
Db 1010 SISLSDQ-----YKQEKILLQRCBETGNAYEDLSQYKAAQ-----1047
QY 754 VKQKALEABELDAQHSETFDNIDLSPKASHRSKQRHJKQSLYGYVDFTRHDDNRSDN 813
Db 1048 -BKNSKLECLLN--EFTSLCENRKELEQLKAPAKHQEFL-----1086
QY 814 FNTGNMTVLSPLYLNTVLPSSSSSGSLDSRSEKDSRLERGERIGLGNYPHATENPGTS 873
Db 1087 -----TKLAPABERNQNLMELETVQALRSEMTDNQNSKSEAG-----1126
QY 874 SKRGL--QISTTAAQIAKMEVBSAHTSQEDRSSGSTTELHCVTDERNALRRSSAATH 931
Db 1127 --GLKQEIWTLKEQNKQKEVNDLLOENEQLMKVMKTKHECONLESEPIRNS-----1177
QY 932 SNTYFTKSENENRITCSM-PYAKLEYKRSNDSLSNSVSDNGYKXGQMKPSIES----Y 986
Db 1178 -----VKRESERNQCNFPQMDLEVKEITSLDSYNA-----QVLQLEAVLRNKLKL 1224
QY 987 SEDDESKFCYQYPADLAHKHS-----ANHMDD-----NDGELDTPIFINSKY 1031
Db 1225 QSEKEKEC-----LQHELOTIRGDLTSLNLDQWQSEISGLKDCBID-----1267
QY 1032 SBEQLNSG--ROSPQNERWAPKHI-----IEBEIKQSEORQSENQSTTYPVYTESTD 1083
Db 1268 ABEKYISGPHELTSDQNDN-----AHLQCSLOTMMKNELEXICETILOQAEKELYTELND 1323
QY 1084 DXHLKFPQFGQOQCVSPYRSRGANGSETNRVCSNHNQVNSQSLCOQDEDDYEDDKPTNY 1143
Db 1324 SR-----SECITATEK-----MAEEVKGULNEVYLINDJSLH 1357
QY 1144 SEYSEEEQHEERPTNYSIKYNEKRHVODPIDVSLKYATDIPSSQKQSFQK-SSS 1202
Db 1358 GELVEDIPGGEFGEQF-----NEQHPVSLAPLDESNSYEHLLTSDKEVQMHFAELQEK 1410
QY 1203 QOSKTEH-----MSSSENTTTPSSNAKRNQLHPSSAQSRSGQ-----1242

Db 1411 FLSSLOHEKHILHDOHCQSSKMSSELOTYYVDSLKAENLVLSTNLRNFGQDLVKEMQLGLEE 1470
QY 1243 ---PQKAATC-----KVSSINQETIOTYCVEDTPICFSRCSLSSLSAEDIEGNCQTTQ 1294
Db 1471 GLVPSLSSCVPDSSSLSGDSSFYRALLEQT-----GMSLLSNLEGAVANQCSV 1523
QY 1295 EADSANTLQ--IAEIKGKTGRSAE-----DPVSE-----VPAYSQHPRTSSRL 1337
Db 1524 DEVFCSLOTYYVDSLKAENLVLSTNLRNFGQDLVKEMQLGLEEGLVPSLSSCVPDSSSL 1583
QY 1338 QGSSLSSESARHKAVE-----FPSGAKSPSKSGAOTPKSP 1372
Db 1584 --SSLGDSFYRALLQGTGDMSSLNLEGVVSANQCSDEVFCSLSQENLTKRTPSAP 1641
QY 1373 PEHYVQETPLMFRCTSVSSL-DSFESRSIASSVQSEPCSGMVSGIISPSDDLDPSPQOTM 1431
Db 1642 AKGVSELESCEVYROSLEKLEKESQGMKNKEIQELEQLLS-----SERQEL 1691
QY 1432 PPSRSKTPPPPTAQ--TKREVPKNKAPTAEKRESGPKQAAVNAAVORQVLP-----DAD 1486
Db 1692 DCLRKQVLSNEQWQKLTSTVLEMESKLAEEKQT--BOLSLEVAKLQLOGLDSSR 1749
QY 1487 TLLHFATESTPDPFCSSSLSALSALDEPFTQDKDVELRIMPPVQVQENDNGNTESEQPK-ES 1545
Db 1750 SLLGIDTDAIQGNESCDIS-----KEHTSETTERTPKHV 1786
QY 1546 NEHQEKAETIDSEKDLDDDDDIIELECIISAMP-TKSSRKGKPAOTASKLPPP 1604
Db 1787 HQICDKDAQQ--DLNLDIEKITETGAVKPTGEGSGQSPDTNYPEPGECKTQGSSECE 1844
QY 1605 VA-RKPSQLPVYKLLPSQNRLOP-QKHVSFTPDGDMPRVYCVEGTPIINFSTATSLDLTI 1662
Db 1845 LSFSGPNALVPMDFLGNQEDIHNLQRLVKETSNENRLHLHVED-----RDRKV 1893
QY 1663 ESPNEL-----AAGEVRRGGAQSGE-----F 1684
Db 1894 ESLNENKELDSLHLQEOVLMTKIEACIELEKIVGELKXKENSDELGEKLYFCDHQELL 1953
QY 1685 EKRTIPTEGRST-DAGQGTSSVTIPELDDNKAEEGDILAE---CINSAMPKGS---1737
Db 1954 QRVET--SEGUNSDLEMHADKSR--EDICDNVAKVNDWKERFLDVENLSIRSEKA 2008
QY 1738 ---HKPRVKKIMDQVQ-----QASASSAPNKNQLDGK---KKK 1771
Db 2009 SIEHEALYLEADLEVVOQTEKLCLEKONENKQKIVIVCLEBELSVVTSERNQIRGELDTMSK 2068
QY 1772 PTPSVKPIPQNTERTVRKRNADSKNNLNAERFSDNKSQKNNKNSKDFNDKLPNE 1831
Db 2069 KTTALDQLSKEMKEKTOELESBOSE-CLHCIOVAEAEVKEKTELLQTLSSDVSLLK---2124
QY 1832 DRVRSFAPDSPHYTPTEGTPTCYFSRNDSSLSLDFDDDDVDSREKAEALRKAKENKSE 1891
Db 2125 -----DKTH-----LQEKLOSLKDSQALSCLKELNQLAQLNKEKE 2162
QY 1892 AKVTSHTEL-----TSNOQSANKTOAIKAPINRGQPKPILOKQSTPQSSKQIDPRG--1944
Db 2163 LLVKESELQARLSESDYEKLVNSKALEAALVEKGE--FALRJSST--QBEVHOLRRGIE 2218
QY 1945 ---AATDEKQLONFAIENTPVCFPSHNSL--SSLSIDIDQE--NNKENEP--IKETEPPDS 1995
Db 2219 KLRVRIEADKQHLIAEKLKERENDSLKDKVENLERELQNEENQELVILDAENSKA 2278
QY 1996 QCEPSPKQOAGYAP--KSFHVETPVCFSRNSSLSSISDSDDL--LQECISS--AMPK 2049
Db 2279 EVETLKTIQIEEMARSLKVFEL-DLVTLRSEKENLTKIQEKQOGLSDELKLLSFKSLLE 2337
QY 2050 KKKPSRLKGDNKEHSPRNMGGILGEDLTLDKQFORPDSEHGLSPDSENFDWKAIOEGAN 2109
Db 2338 EKEQAEIQIEESKATVEM-----LONQIKELN--EAVNALCGDQEIM--KATEQSLD 2386
QY 2110 SIVSSLHQAAA-----ACLSRQASSDSLSLSKSGI-SIGSPFHLPD 2153

Db 2387 PPIEEHQLRNSIEKLRLARLEADEKKQLCVLQQLKESEHDLKGRVENLRELEIART 2446
Qy 2154 QEEKPFSTNGKPRILKGEKSTLTETKIESEKIKGKKYKSLITGKVRNSIEGQM 2213
Db 2447 NOEHAALAEANSK-----GEVETLKA-KIEGWTQSLRG---LELDVVT--IRSEKE---DL 2493
Qy 2214 KQPLQANMPSISRGRTMIHPIGVNNSSTSPVSKGPPKLTFAKSPSGQATTSPRG 2273
Db 2494 TNELOKQERISELE-----IINSPENILQKEQKVMKEKST----- 2534
Qy 2274 AKPSVKSELSVAROTQIOGSSKAPRSRSTPSRPAQPLSPRIQSPGRNSISPR 2333
Db 2535 AMEVLQTLQKELNERVAAL-----HNDQEAACKAEQNLSSQVECLEKA----- 2579
Qy 2334 NGISPPNKLSPRTSPSTASTKSSGKMSYTSPPGROMSQNLTKQTGLSKNASSIPR 2393
Db 2580 -----QLQGLDEAKNNIVLQSSVNGLIQVEDGKQLEK---KDEISRLKNQIQD 2629
Qy 2394 SESASKGLNQV-----NNGANKVKVELSRMSTKSSGSEDRSERPVLVRQSTFI 2444
Db 2630 QEQLVSKLSQVEGBHQLWKEQNLRLNLTVELEQKIQVLQSKNASLQDTLEVL--QSSY- 2686
Qy 2445 KEASPTLRKLESASFESLSRSSRSPASPTRSQATPVLSPDLMSLSTHSSVQ--AG 2502
Db 2687 -----KNLENEBELTKMDKMSFVEKYNKMTAKETELQREMHMAQKTAELOBELSG 2737
Qy 2503 GWRKLPNLSPTIEYNDGRPAKRDHDIARSHSESPSRL-----PINRSGTWKRE----- 2550
Db 2738 EKRELQELQLLLEIKSSQDLKELTLENSLKKSLDCHMKDOVEKGVREIAYQL 2797
Qy 2551 -----HSHKSSSL-----PRVSTWR-RTGSSSSILSSSSSSSKAKSEDEKHNVSIS 2596
Db 2798 RLHEAEKKHQALLDNTNKQYEVETIQTREKLTSEKLECLSSQKLEIDLKSSKEELNLSLK 2857
Qy 2597 GTKQ-----SKENQVSAKGTW-----RKIKENEFSPIN 2624
Db 2858 ATTQILBELKTYMDNLKYNQLKXENE-RAQGNKMLIKSKQLBEEKEILQKLSQLQ 2916
Qy 2625 STGQTVSSGA--TNGAESKTLIYQMAPV--SKTEDVWV-----RIED----- 2663
Db 2917 AAQEKQKTGTMTDKVDELATEIKELKTELEKTEADEVLDKYCSLLISHEKLEKAKEM 2976
Qy 2664 -----CFINNPRSGR-----SPTGNTPPVIDSVSEK--ANPNTKDSKDNQAKQNV-- 2706
Db 2977 LETQVHLCSQQQKDSRGSGPLGVPVGPSPISVTEKRLSSGQNKASKGKQRSGIWE 3036
Qy 2707 -GNGSVPMRTVGLNRLTSPFIQV---DAPQKGTETKPGQNNP-----VPVSETNE 2753
Db 3037 NGGQPTPATPESFSKSKKAVMSGIHPAEDTEGTEFEP-EGLPVVKKGFDIPTGKT-- 3093
Qy 2754 SP-IVERTPSSSSSKHSPSGTVAAR---VTPFNYNPSRKSADSTS--ARPSQIP- 2806
Db 3094 SPYILRETTWATRTSPR-----LAAQKALSPLSLGKENTLAEBSKPTAGGSRQKVKV 3146
Qy 2807 ---TPVNNNTKXDSKTDST-----ESSGTQSPK 2832
Db 3147 AQRSPVDSGTILREPTTKSVPVNLPERSPTDSFR 3181

RESULT 14
MAPB RAT

ID MAPB RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9BR21; Q9OW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

RN
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN-Sprague-Dawley; TISSUE-Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenchingloh G., Hermans-Borgmeyer I., Kirsch J.,
Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nohias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC
CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC
CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC
CC -!- INDUCTION: By nerve growth factor.
CC
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC
CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC
CC -!- PTM: Phosphorylated.
CC
CC -!- SIMILARITY: TO MAP1A.
CC
CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; U52950; AAB17068.1; -
DR EMBL; X60370; CAC16162.1; -
DR EMBL; X16623; CAA34620.1; AUT_SEQ.
DR PIR; A56577; A56577.

RESULT 15
 ID ANK3_HUMAN STANDARD; PRT; 4377 AA.
 AC Q12955;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ankyrin 3 (ANK-3) (Ankyrin G).
 GN ANK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=95138209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.;
 RT "Ankyring. A new ankyrin gene with neural-specific isoforms localized
 RL J. Biol. Chem. 270:2352-2359(1995).
 CC -!- FUNCTION: Membrane-cytoskeleton linker. The neural-specific
 CC isoforms may participate in the maintenance/targeting of ion
 CC channels and cell adhesion molecules at the nodes of Ranvier and
 CC axonal initial segments.
 CC -!- SUBUNIT: Neural-specific isoforms may be a constituent of a
 CC neurofascin/NRCAM/ankyrin G complex.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=A number of isoforms are produced;
 CC Name=1; Synonyms=480-kDa isoform;
 CC IsoId=Q12955-1; Sequence=displayed;
 CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues. Isoform
 CC 1 is neural-specific.
 CC -!- SIMILARITY: Contains 23 ANK repeats.
 CC -!- SIMILARITY: Contains 1 death domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U13616; AAA64834.1; --
 CC PIR; A55575; A55575.
 CC HSP; P55273; IBI8.
 CC Genew; HGNC:494; ANK3.
 CC MIM; 600465; --
 CC GO; GO:0006605; P:protein targeting; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000489; Death.
 CC InterPro; IPR000906; ZU5.
 CC Pfam; PF00023; ank; 24.
 CC Pfam; PF00531; death; 1.
 CC Pfam; PF00791; ZU5; 1.
 CC PRINTS; PR01415; ANKYRIN.
 CC SMART; SM00248; ANK; 21.
 CC SMART; SM00005; DEATH; 1.
 CC SMART; SM00218; ZU5; 1.
 CC DR PROSITE; PS50089; ANK_REPEAT; 21.
 CC DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC DR PROSITE; PS50017; DEATH_DOMAIN; 1.
 CC Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
 KW REPEAT 73 102 ANK 1.
 FT REPEAT 106 135 ANK 2.
 FT REPEAT 139 168 ANK 3.
 FT REPEAT 172 201 ANK 4.
 FT REPEAT 203 230 ANK 5.
 FT REPEAT 234 263 ANK 6.
 FT REPEAT 267 296 ANK 7.
 FT REPEAT 300 329 ANK 8.

FT REPEAT 333 362 ANK 9.
 FT REPEAT 366 395 ANK 10.
 FT REPEAT 399 428 ANK 11.
 FT REPEAT 432 461 ANK 12.
 FT REPEAT 465 494 ANK 13.
 FT REPEAT 498 527 ANK 14.
 FT REPEAT 531 560 ANK 15.
 FT REPEAT 564 593 ANK 16.
 FT REPEAT 597 626 ANK 17.
 FT REPEAT 630 659 ANK 18.
 FT REPEAT 663 692 ANK 19.
 FT REPEAT 696 725 ANK 20.
 FT REPEAT 729 758 ANK 21.
 FT REPEAT 762 791 ANK 22.
 FT REPEAT 795 825 ANK 23.
 FT DOMAIN 1519 1898 SER-RICH.
 FT DOMAIN 4090 4174 DEATH.
 SQ SEQUENCE 4377 AA; 480399 MW; F42379E55768B684 CRC64;
 Query Match 2.6%; Score 380.5; DB 1; Length 4377;
 Best Local Similarity 18.3%; Pred. No. 1.4e-06;
 Matches 595; Conservative 430; Mismatches 1178; Indels 1051; Gaps 147;
 QY 39 TEASNMKEVLKQLOQSIEDAMASSGOIDLLERLKEINLDSSNPPGVKLRSMKMS-----92
 DB 1695 TWASSLSFVKQMPG--HAELVNGSISPLKYASSSTLINGCKATATLOEKISSATNSV 1752
 QY 93 LRSYSGREGSV-----SSRSGEC 110
 DB 1753 SSVWSAATDTVEKVFSTTTAMPFPLSYSAAPSAFQSLRTPSASALYTSLSGSSISATT 1812
 QY 111 SPVPMG--SPRRGFVNGSRSTCYLELEKERSLLADLDKEKEKDYWYAOQLNTKR 168
 DB 1813 SSVTSIIITPVYVNVNLPPEA-----LKK-----LPDSNFTKSAALLSPIKLTUTE 1862
 QY 169 IDSFLTENFSLQDLTRRQLEYEARQIRVAMEPQLGTCQDMKRAQRRIARIQIIEKDI 228
 DB 1863 THQPQ---HPSRTSPVKSSLFLAPSALKLSTPSSLSSQEIILK-----DVAEMKEDL 1912
 QY 229 LRIQLLOSQATEERSQNKHETGSHDAER-----QNEGGVGEI---NWAISNG 277
 DB 1913 NMRTAILQTDVPEKPFQPELPKEGRIDDEEPFKIVEKVEDLVKVEIILKDVCDVNDKG 1972
 QY 278 QGSTTRMD-----HETASVLSSTSHSAPRRLTSHLTGKVENVYLLS 320
 DB 1973 SPKSPKDKGHSPEDDWIEFSSEIREARQQAASQSPSLPERVQVKAASEKDYNLTK 2032
 QY 321 MCGTHDKDDMSRTLLAMSSQDSCISMRQSG---CLPLLIQLLHGNKDSVLLGNRGS 376
 DB 2033 VIDYLTNDIGSSSLTNLYKPFDAKDGEGQKRVLPALQEHKLKMPASPMTSTSE 2092
 QY 377 KEARARASAL-HNIHSQPDKRGREIRVLHLEQIRAYCETCWEWQEAHEPGMDQDK 435
 DB 2093 KELCNADSPFGTDTILESDDD-----FSQH-----DQDK 2122
 QY 436 NPM-----PAPVEHQICPAVCVLMKLSFDEEHRHAMELGL 472
 DB 2123 SPLSDSGFETRSEKTPSAPOSAETGPKPLFHEV-PIPPVITETRTVEVH-----2171
 QY 473 QAIARELLQVDCMYGLNDHYSLTREVAGWA-----LTNLTGADVANKAT 518
 DB 2172 -----VIRSYDPSAGDVPTQPEEPVSPKPSPTTFMELEPKPT 2208
 QY 519 LCSMKGMRALVAQLKSESEDLOQVIASVLRNLSWRADVNSKKTLEVGSKALMECALE 578
 DB 2209 TSSKEKVKAFQMKASSEEDHNRVLSKGMK-----2239
 QY 579 VKKESLTKSVLSALWNSAHTENKADICAVDGAFLVGLTYRSQTNLTALIESGGGI 638
 DB 2240 VKKET-----HIT-----TTRMYVTHSGPG- 2259
 QY 639 LRNVSSLIATNEDHRQILRENNCLQTLLOHLKSHSLTIVSNACQTLWNLSARNP-KDQEA 697

Db 2260 -----GEGASERIEETMSVHDMKAFQ-----SGRDPSELAK 2292
Qy 698 LWD-MGAVSMLKNLHSHKMIAMGSAALNLANRPAKYKDANIMSPGSSLSPLHVRK 756
Db 2293 LFEHKSAYS-----PDVHKSAETSQAHA-----KD-NQWKP-----KL 2326
Qy 757 OKALEAEILDAOHLSETFNIDNLSFKASHRSKORHKOSLYGDYVDFNRHDDNRSNFT 816
Db 2327 ERIIBVHEKNOAEPTEVI-----IRETKKHPEKEMVYQKLSRGDNLKD----- 2374
Qy 817 GNMVLSPLNTVTLVPSSSSRGSLDSRSRSEKDRSLERERIGLGNVHPATENGTSSKR 876
Db 2375 -----FLPEKHDA-----FPCEEQOQEE 2395
Qy 877 GLQISTTAQIAKVMEEVSAIHT-----SOEDRSGSTTELCVTDERNALRRSAAHT-----H 931
Db 2396 EL-----TAESLPSVLESRVNTPVQEDSRPSSAQL-----ISDSDYKTLKLLSHQSIYEH 2449
Qy 932 SNTYNFTKSNRNTCSMPYAKLEYKRSNDSLSNSVNDGYK-----RGQWKPIESYSE 988
Db 2450 DDELSERLGE-SYRFAERMLLSEKLDVSHSDTEESVTDHAGPPSSSELOQSGDKRSREKIAT 2508
Qy 989 DDESKFCS--YGOYFADLAHKHSANHMDNDGELDTIPNYS-----LKYSDQ 1035
Db 2509 APKKEILKIYDVSNGVGVKSKDEHFK-----VTVLHSGNVSSPKHAMWRFTEDR 2563
Qy 1036 LNSGRQSPSONERWAPKHIIEDEIKQSEQRQSRNQSTTYPYTESTDDKHLKTQPHFGQ 1095
Db 2564 LDRGREKLIYEDR-----VDRTVKEAEKLT-EVSOFFRDKTEKLAND----- 2604
Qy 1096 QECVSP-VRSRGANGSEINRVGNSHGINQVQSQCQEDDDYEDDKPTNYS----- 1144
Db 2605 -ELQPEKKARPKNGKE-----YSSQOPTSSSEKULLTELL 2640
Qy 1145 -----EERPTNYSIKYNEERHVDQPTDYLSKYATDIPSSQ 1192
Db 2641 ASNDEWVKARQHPDQGPFAEKAP---SLPSPPEXWMLSQOTEDS-KSTVEAKGSGS 2696
Qy 1193 QSFPSKSSGSSKTEHMS-----SSENTSTPSSNAKQNLHPSQASGQPOKKAAT 1248
Db 2697 QSKAPDGGQFQKQSKLSIRLFEQGTAKSKDMQEDR--KSDQSR----- 2745
Qy 1249 CKVSSINOTIQTVCYEDTPICFSCSSLSLSAEDIEGNC-----QTTQEADSANTL 1302
Db 2746 IPVKIOESKLPYQV-----FAEKQOKADLDPDESUSVQKDFWMLTKDEHAQSNEL 2799
Qy 1303 QIAEIKGKTGTRSAEDPVSEPAVSGHPRTK--SRLOGSSLSSESARHKAVEFPG--GAK 1359
Db 2800 VVND-----SGSDNVKK-----QTEMSSKAMPDSFSEQAKOLACHITSDLATR 2844
Qy 1360 SP-----SKSGAOTPKSPE-----HYVQETPLMFRC'TSVSSLDSFESRSIAS 1403
Db 2845 GPWDKKVFTWSSGATNNKSKQEKLSHVLVHDVRENHIGHPESKSVQOKNEF--MSVTE 2902
Qy 1404 SVQSEPCSGMWGSI-----ISPSDL-----PDSPGOTM-PPSR-----SKTPPPPP 1443
Db 2903 RERKLLTNGLSLSHIEKMTVKFSKKVLYREYVVKEGDHPGGLDQPSRRSESSAVSHIPV 2962
Qy 1444 QTAQTKREYVKNKAPTAERKESGPKQAAVNAVQVLPDADTLHFAATESPPGFSCS 1503
Db 2963 RVADERMLSSNIPDGFCEQSPAPPHLSOKLSQSSMSKETVET-QHF--NSEDIEKVTY 3019
Qy 1504 SLSALSLEPFTOKDVELRIMPVQENDNGNETESEQPKESNENOEKAEKTIDSEKDL 1563
Db 3020 SEISKVSKQSYV-----GLCPLEETET-SPTKSPDSLSFSPGKESPS-----SDV 3065
Qy 1564 LDSDDDDDIIELEECIISAMPTKSKRGKKPAQATASKLPPVA-----RKSQLPV 1614
Db 3066 FDHSPIDGLEKL-----APLAQTEGKE-----IKTLPVYVSFVQVQKYEKIQOQGV 3114
Qy 1615 YKLLPSQNLQPKHVSFTPGDDMPVYCVETGP-----INFSTATSLDLTIESPEN 1667
Db 3115 KLIISOECKTVQETRGTFYTTROQKPPQSPGSPEDDTLEQVSLDSSGKSPITPFPSS 3174

Qy 1668 ELAAGEVRGGAQSGEFE-----KRDPTTEGRSTDEAQQGKTSVITPELD 1714
Db 3175 E-----EVSYETSKTDPDSLIAYIPGKPSPIEVSESEEEQAKSTS-----LK 3219
Qy 1715 DNKAEBEDILAEINCAMPKGSHPK--PRVKKI-----MDQVQOASASSAPNK- 1762
Db 3220 QTTVEETAVEREMPND-VSKDSNQRPKNRVAIEBPFPPLDADQIESDKHHYLPK 3278
Qy 1763 -----NQLDGKKK-----KPTSPVKP--IPONTEYRTRVRKQADSKNNNAER 1803
Db 3279 VDMIEVNLQDEHDKYQLAEPVIRVQPPSPVPPGADVSDSDDESIVQPVVKYTFKLKE 3338
Qy 1804 VFSNKKSKQKNUK--NNSKDF--NDKLPNNEDVRGSPAFDSPHYTIEG----- 1851
Db 3339 VDDEQEKPKASAEKASNOKELESNGSKDNE-----FGLGLDSPQNEIAQGNNDQSITE 3394
Qy 1852 ----TPVCFSEN-----DSLSSLDFFDDVDLSREKAELR-KAKENKSEAKVTGHTEL 1900
Db 3395 CSTATTABRSHDTEIDSLDGYDLODEDDGLTESDKLPICAMEIKKDIWTEGILKP 3454
Qy 1901 TSNQOQANKTQAIKQPINRGQPKPILOKQSTPPQSSKDIIDRGAATDEK--LQNFAREN 1958
Db 3455 ADSPSQSKLEVIEE-----GKVGPDDEKPPSKSSSEKTPDK---TDQKSGAQFFTEG 3507
Qy 1959 -----TPVCFSHNSLSLSLDIDQENNNKENEP-----KE 1989
Db 3508 RHPDRSVFPPTYSYKVDEEFATPFTVATKGL---DFDPWNNRGDDDEVFDSKREDE 3563
Qy 1990 TEP-----PDSQGEPSKPAQGYAPKSFHVEDTPVCFSRNSLS---SLSID 2033
Db 3564 TKPFLAVERDRSPATTPDT--TPARTDDESTPTS--BNPFPFHEGKMFEMTRSGAID 3618
Qy 2034 -SEDDLLQ-----CISAMPKPKSRLLKGDNE--KHSRNMGGILGEDTLIDL---KD 2082
Db 3619 MSKRDVFEERLQFPQGEHTSEKSGQGGKSMVTATPQPSG---DTTVEITNLERN 3674
Qy 2083 IQRPDPSHGLS-PDSENFDMKAIQEGANSIVSSIHQAAAAACLSROASSDSDSILSKSG 2141
Db 3675 VEPTVEPNPSITSGE---COEGTSS--SGSLEKSAATNTSK---VDPKLTPIKMG 3725
Qy 2142 ISLGS-----PFHLTPDQEPKFPFTSNGK-----PRILKPGK----- 2173
Db 3726 ISASTMTWKKEGPEIT-DKIEAVMTSCQGLENETITMISNTANSQMGVPRHEKHDFOKD 3784
Qy 2174 -----STLETKKIES-----ESKIGKGGKKVYKSLITGK--VRS 2205
Db 3785 NFNNNNLDSSTIQDINIMSNIVLTHESAFTCTEKDNPKVSSGKT--GVLQGHCVRD 3842
Qy 2206 NSEISGOMKQPLQANMPSISRGRTMIHIPVRNSS-----SSTGPVSKGPP---LKT PAS 2258
Db 3843 KQKVLGEQOKTE-----LIGIROKSKLPKATSP-KDTFFPNHMSNTKAS 3887
Qy 2259 KSPSEGQATATSPRCAKPSVKSELSPVARQTSQIGSSKAPSRSGSRDSTPSRPAQPLS 2318
Db 3888 KMKQVSO-----SEKALTTSVCVDKSRIPVKNTPRDNIA----- 3925
Qy 2319 RPIQSPGRNSISPGRNGISPPNKLSQLP-----RTSSPSTASTKSGSGKMS 2365
Db 3926 -----VRKACATQKQOPEKGAQKLPKLPVKVRSTCVTTTTTTTTTTTTTTTT 3979
Qy 2366 YTSRGRMSQONLTKOT-----GLS-KNASSIPR-SESASGLNQMNANGANKKVELSR 2418
Db 3980 CTVKVRSQLKEVCCKHSIEYFKGISGETLKLVDLSSEKMOSELS-----DEEE 4030
Qy 2419 MSSTKSGSEDRSERPVLVRQSTFIKEAPSPTRRLRRLKESASFESISPSRRPASP--TRS 2477
Db 4031 STSNTLSLSEISRGQSVITKSRDKKTEAPLKSSEKAGEKRSRRTPGQSPCERT 4090
Qy 2478 QAQTPVLSPLPMSLTHSSVQAGGWRKLPNPIPTIEYNDGPAPKRDHIARSHSPSS 2537
Db 4091 DIRMAIVADHL---GLS-----WTELARELNFVSD-----EINQIRVENPN 4128

```

QY 2538 RLPINRS-----GTWKREHSHSSSLPRVSTWRRTGSSSSILSASSESESEKAKSEDE-KHV 2592
Db 4129 SL-ISQSFMLKKWTRDGNAT-----TDALTSVLTIKINRIDIVITLLEGPFDY 4177
QY 2593 NSISGTKO-SKENQVSAKGTWRKIKENEFSTNSTSQTVSSGATNGAESKTLIYQMAPAV 2651
Db 4178 GNISTRSFADENNVE-----HDPVDMQNETSG-----4207
QY 2652 SKTEDVVRIEDCPINNPRSGRSPGTPTPPVIDSVSEKANPNIKDKDNQAKQNVGNSV 2711
Db 4208 -----NLESC-----AQARRVTGGLDLRLDDSPQCRDSTSYLKGAGKFEANGS- 4253
QY 2712 PMRTVGLLENLTFSIQVDAPDKGTEIKPGONNPPVSVETNES--PIVERTPFSSSSSSK 2769
Db 4254 -----HTEIT-----PEAKTSYFPBSQNDVGKQSTKETIKPKIH-----GSGH 4292
QY 2770 HSPFSGTVAARVTPFNYPNPKSSADSTS-----ARPSQIPTVNNNTKKRDSKTDST 2823
Db 4293 VEEFASPLAA-----YOKSLEETSKLIIETKPC---VPVSMKXMSRTSPADGK 4338
QY 2824 ESSCTQSPKRHSGS 2837
Db 4339 PRLSHEEGSSGS 4352

```

Search completed: August 25, 2004, 17:16:32
Job time : 76.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:05:26 ; Search time 183.5 Seconds
(without alignments)
4888.383 Million cell updates/sec

Title: US-09-442-489F-2
Perfect score: 14575
Sequence: 1 MAASYPQLLKQVAKWEN.....ESSGTQSPKRHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10530	72.2	2829	13 P70039	P70039 xenopus lae
2	5156	35.4	1056	11 Q8BNP7	Q8BNP7 mus musculus
3	3508.5	24.1	2274	11 Q921K7	Q921K7 mus musculus
4	3414.5	23.4	2303	4 Q95996	Q95996 homo sapien
5	2229	15.3	489	11 Q8BRD8	Q8BRD8 mus musculus
6	2172.5	14.9	733	4 Q9UBZ1	Q9UBZ1 homo sapien
7	1989.5	13.7	1246	4 Q9Y632	Q9Y632 homo sapien
8	1778.5	12.2	2416	5 P91667	P91667 drosophila
9	1763	12.1	2417	5 Q9VAS9	Q9VAS9 drosophila
10	1705	11.7	1685	4 Q9UEM8	Q9UEM8 homo sapien
11	1249	8.6	324	11 Q8C493	Q8C493 mus musculus
12	1221.5	8.4	1067	5 Q9VIT2	Q9VIT2 drosophila
13	1220.5	8.4	1067	5 Q961B0	Q961B0 drosophila
14	1041	7.1	208	4 Q9P119	Q9P119 homo sapien
15	756	5.2	159	4 Q722Q8	Q722Q8 homo sapien
16	680	4.7	146	11 Q8C9I9	Q8C9I9 mus musculus

ALIGNMENTS

RESULT 1

P70039 PRELIMINARY; PRT; 2829 AA.

```

ID   P70039          PRELIMINARY;          PRT; 2829 AA.
AC   P70039; P79934;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Adenomatous polyposis coli.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC   Xenopodinae; Xenopus.
CX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Vlemnickx K., Wong B., Guger K., Gumbiner B.M.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DBSJ databases.
DR   EMBL; U64442; AAB41671.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR000225; Armadillo.
DR   Pfam; PF00514; Armadillo_seg; 4.
DR   SMART; SM00185; ARM; 5.
SQ   SEQUENCE 2829 AA; 310878 MW; 8A2BABDB7706E496 CRC64;

```

Query Match 72.2%; Score 10530; DB 13; Length 2829;
Best Local Similarity 74.3%; Pred. No. 0;
Matches 2128; Conservative 263; Mismatches 416; Indels 58; Gaps 40;

```

QY   1 MAASYPQLLKQVAKWENSLRQELDSSNHLTKLETSANWKEVYLKOLQGSIEDAM 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   1 MAASYPQLKQVAKWENSLRQELDSSNHLTKLETSANWKEVYLKOLQGSIEDAM 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   61 ASSQIDLLERLKLNLDSNFFGVKLRKMSLRSGVSSRSGECSPVPMGSPFR 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   61 ASSGPIDLLERLKLNLDSNFFGVKLRKMSLRSGVSSRSGECSPVPMGSPFR 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   121 RGFVNGRSTCYLELEKEKRSLLADLKEEKKWYLAQLNLTGKIDSLPTENFSL 180
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   121 RGLNGRSAGYMELEKEKRSLLADLKEEKKWYLAQLNLTGKIDSLPTENFSL 180
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


181 QDLTRQLEAQRIRVAMEEQLGTQODMEKRAQRIRIQTOKDILRIQLLOSQAT 240
181 QDLTRQLEAQRIRVAMEEQLGTQODMEKRVQTRVGKIHOIEBELLAIQILLOSQVA 240
241 E-ABRSQNKHETGSHDAERQNEQGVCEINMATS-NGQGSTTRMDHETASVLSSTH 298
241 EABERTTQSHDAGSRDAEKLPGQGGTSETAGNVGSGGSSRADHDHTTVMNSNTY 300
299 SAPRLTSHLGTKEYMYVYLLSMGLTGDHDDMSRTILAMSSQDSCISMESQGLPLLIQ 358
301 SVPRLTSHLGTKEYMYVYLLSMGLTGDHDDMSRTILAMSSQDSCILAMEQSGCLPLLIQ 360
359 LLHGNDDKSVLLGNRSRGEKABARASAAALHNI THSQPDDKRGREIRVLHLLBOIRAYCE 418
361 LLHGNDDKSVLLGNRSRGEKARASGAALDNI THSQPDDKRGREIRVLHLLBOIRAYCE 420
419 TCWEQZAHFPGMDQKNPMPAPVEHOI CPAVCVLMKLSFDEBHRHAMELGGLOIAEL 478
421 TCWEQZAHFPGMDQKNPMPAPVDHOI CPAVCVLMKLSFDEBHRHAMELGGLOIAEL 480
479 LOVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMSKGMRALVAQLKSESE 538
481 LOVDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMSKGMRALVAQLKSESE 540
539 DLOQVIASVLRNLSWRADVNSKTLREVGSVKALMECALVEKKESTLKSVLKALWLSAH 598
541 DLOQVIASVLRNLSWRADVNSKTLREVGSVKALMECALVEKKESTLKSVLKALWLSAH 600
599 CTENKADICAVDGALEFLVGLTLYRSQNTLAIIESGGGILRNVSLLIATNEDHROILRE 658
601 CTENKADICAVDGALEFLVGLTLYRSQNTLAIIESGGGILRNVSLLIATNEDHROILRE 660
659 NNCLOTLOHLKSHSLTVSNACGLTNLSARPKDOEALWMDGAVSMKLIHSHKMI 718
661 NNCLOTLOHLKSHSLTVSNACGLTNLSARNAKDOEALWMDGAVSMKLIHSHKMI 720
719 AMGSAALRNLMANRPAYKDANIMSPGSLPSLHVKKQKALEAELDAOHLSETFDNIDN 778
721 AMGSAALRNLMANRPAYKDANIMSPGSLPSLHVKKQKALEAELDAOHLSETFDNIDN 780
779 LSPKASHRSQRIKOSLYGVDFDTNHDN--RSDNFTNGMNTVLSPLYNTTIVLPSSSS 836
781 LSPKTTTRNKQRKQNLCSYALDSSRHDDSI CRSDNFSIGNLTVLSPLYNTTIVLPSSSS 840
837 SRGSLDSRSEKRSLEREGIGLGNVHPATENPGTSSKR-GLQISTTAAQIAKVMEEVS 895
841 PRPTMDSRPEKD----RERTAGLGNVHSTTESGNSKRIQIQLSTT-AQISKVMDVVS 895
896 ALHTSQDRSGSTTELHCVTDERNALRRSSAAHTSNNTYNTFTKSNRNTCSMPYAKLE 955
896 NIHLVQENRSGSGASEMHCMDSERNRQKPSNHPQSNPFTFTKARSSTRGCPVAFMKME 955
956 YKSSNDLSNVSNDGYSRGKQKPSIESYSDDESKTCSGOYPADIAKHTHSANWMD 1015
956 YKASNDLSNVSSTEGIGRQGVKQSVESYSDDESKTCSGOYPADIAKHTHSANWMD 1015
1016 DNGELDTPIYSLKYSDEQLNSGRSPQNERWARPKHIIIEDEIKOSEQCRSNOSTTY 1075
1016 DNDTELDTPIYSLKYSDEQLNSGRSPQNERWARPKHIIIEDEIKOSEQCRPRTKTTY 1075
1076 PVYTESDDKHLXFOHFGQOECVSPY-RSRGANGS-EINVEGSHGINQVNSQLCOED 1133
1076 SSTYENKEKHKKPPPHFNQSENVNAYTRSRGANNQVQDSRVSSNLSNNSKASKPHCQVD 1135
1134 DYDDKPTNYSERYSEBEEQHEEB-ERPTNYSIK-VNEEKRHVDQPIDYSLKVATDIPSS- 1190
1136 DYDDKTTNFSERYSEBEEQDETERQNKYNIKAYASEEHGQPIDYSEKYSTVDVPSSA 1195
1191 QKQSFPSKSSSQSSKTEHMSSESTSTPSSNAKROQLHPSVAQSRG--QPOKAAT 1248
1196 QKPSFPYNNSSKQPKPKQEVSSNS-NTPTPSPNSNRQNLHPNSAQSRPGLNRPKQIN 1254

1249 KCVSSINOETIQTVCVEDTPICFRCSSLSLSSAEDEI-CGNQTTQEADSANTLOIAEI 1307
1255 -KPPSINOETIQTVCVEDTPICFRSGSLSSLSAEDEIEGRNRSRQESNNTLQITEP 1313
1308 KKGITRAEDPVSFPAVSOHPRTKSRLOGLSLS-SESARHKAVRPPSGAKSPSKSGA 1366
1314 K-EISAVSKDGAVENTRSSVHRTTKNRLQTSNISPSDSRHKSVFSSGAKSPSKSGA 1372
1367 QTPKSPPHYVQETPLMFESRCTSVSSSDSPESRSIASSVQSEPCS-GMVSGIISPSDLPD 1425
1373 QTPKSPPHYVQETPLMFESRCTSGSSLDSPESHISASSIVASEHWISGIIISPSDLPD 1432
1426 SPQTMPPSRSKTTPPPPTAQTREVPKVPKAPAEKRESGPKQAANVAQVQVLPDA 1485
1433 SPQTMPPSRSKT-PPPTQVCAKDGSKPIVPEERKVA--KTAVHSAIORVQVLOEA 1489
1486 DTLHLFAESTPDGFSOSSLSALSLEDPFTQKDOVELRIMPPVOENDNGNETESEQPKES 1545
1490 DTLHLFAESTPDGFSOSSLSALSLEDPYIQDVQLKIMPPVLENDQGNKAEFE--KEF 1547
1546 NENOBKEAKETIDSEKLLDDSDDDIEIIEBECIISAMPTKSSRKGKK-PAQTASKULPPP 1604
1548 IDNKAKEDKSEOEKMLDDT-DDDIDIEBECIISAMPRKPSRKNKKVPQTPGKPPPP 1606
1605 VARPSQVPLVYKLLPSQNLQPOKHVSFTPGDDMPRVYCVEGTPIFSTATSLSDLTIES 1664
1607 VARPSQVPLVYKLLSSQNLQTKHVNFTHSDDDMPRVYCVEGTPIFSTATSLSDLTIES 1666
1665 PPNELAAGEYRGGAQSGEFEKRTIPIEGRSTDEAOGKTSVTIPELDDNKAEEGDIL 1724
1667 PPSE-PTNDQNTDLSLTLEKRTIPIEGRSTDDTDASKPLNPT-TVLDEKAEEDIL 1724
1725 AECINSAMPKSKSHKPRVKIMDOVQOASASAPNKNQL-DGKKKKPTSPVKPIPONT 1783
1725 AECIHSAMPKSKSHKPRVKIMQOINHTSAATSGNSRSMQETDKNKPTSVPKPMQSI 1784
1784 EYRVRKRNADSKNLAERVSNDKDSKQNLKXNKGDFNDKLPNNEDRVRGSFAFDSP 1843
1785 GFKERLKNTELKNPSENQYC--DPRKPSKKPSKVANEKIPNNEERTKG-FAFDSP 1840
1844 HHYPIEGTPTCFRNSLSLSDPDDDDVLSREKAEKLRKAKENKSEAKVTSHTLTSN 1903
1841 HHYPIEGTPTCFRNSLSLSDPDDDDIDLSKEAEKLRKEKGTQDDQVKYKHEKRAI 1900
1904 QOSANKTQALAKQPINRGQPKPILOKQSTFPQSSKDIPIDRAAATDEKLQNFATENTPVCF 1963
1901 NPMGKQDQTPKSLGGRDQPKALVQKPTSFSSAAKGTQDRGGATDEKWFATENTPVCF 1960
1964 SHNSLSLSLSDIDQNNKNENEPKETEPPDPSQGEPSKQASGVAPKSFHVEDTTPVCFSR 2023
1961 SRNSLSLSLSDIDQNNKNENEPKETEPLKQGTSETQLGLRRPQTSYVAPKSFHVEDTTPVCFSR 2020
2024 NSSLSLSLSDIDEDLLQECISSAMPKKKPSRLKGDNEKHSRNMGGILGE--DLTLDLK 2081
2021 NSSLSLSLSDIDEDLLQECISSAMPKKKPSKIKNEVCKRSRNSVGGILAEEDPLTDLR 2080
2082 DIQRPDSEHGLSPDSNFDWKAIQEGANSIVSVLHQAAAACLSRQASSDSLSLKSG 2141
2081 DIQSPDSENAFSPDSNFDWKAIQEGANSIVSVLHQAAAAGSLSRQSSDSLSLKSG 2140
2142 ISLGSPPHLLTPDQEKFTNKGPRILKPGKSTLETKTESKGLKGGKVKYKSLITG 2201
2141 ISLGSPPHLLTPDQEKFTNKGPRILKPAKSALENKKEETEEPKGLKGGKVKYKSLITG 2200
2202 KVRNSEISGOMKQPLQANMPSISRGRTMIHIGVRNSSSTSTFVSKKGPPLTPASKSP 2261
2201 KSRSSDFSHCKQSVQTNMPSISRGRTMIHIGVRASSPSTFVSKKGPVFNVPKSGS 2260
2262 SEGOTATISPRGAKPSVKSELSPVARTSOIGSSKAPSGSGRSDTSPSPACQSRPI 2321
2261 NENPSSSSKGTGP-LKSELVYSGRPSSTFGSSKGNRSGSRDSSASSRPSQPJSRPL 2319
2322 QSPGRNISIPGRNISPPNKLSQLPRTSSPSTASTKSSGSGKMSYTSFGRQMSQONLTQK 2381

2320 QSPGRNLSPOKNGISPPNFKSQPRITTSPTASTKSGSGRMVYITSPGQLSQPNLSKQ 2379
Db
2382 TGLSKNASSIPRSESAGKLNQNMNGANKKVELSRMSSTKSGSGSRSRSPVLVRQS 2441
Qy
2386 SGLPKTHSSI PRSESASKLNQ- NVNTGSKKVELSRMSSTKSGSGSRSRSPALVRQS 2438
Db
2442 TFIKEAPSPILRRKLEESASPSISPSRSPASPTESQAOTPVLSPLDMSLTHSSVOA 2501
Qy
2439 TFIKEAPSPILRRKLEESASPSISPSRSPASPTESQAOTPVLSPLDMSLTHSSVOA 2497
Db
2502 GWRKLPNLSPITTEYNDGRPAKHDIAARSHSESPSRLPINRSQTKWREHSKHSLSLPRV 2561
Qy
2498 GWRKLPNLSPITTEYNDGRPAKHDIAARSHSESPSRLPINRSQTKWREHSKHSLSLPRV 2555
Db
2562 STWRRTGSSSILSASSSESSEKAKSEDEK-HVNSISGTYKQSKENQVSAKGTWRKIKENEF 2620
Qy
2556 STWRRTGSSSILSASSSESSEKAKSEDEKQVCSFPQPR--SECSSSAKGTWRKIKSEI 2613
Db
2621 --SPNTSQTQVSSGATNGAESTLIIYQMAVASKTEDVWVRIEDCPINPRSGRSPGTN 2678
Qy
2614 LETPSNGSSSIAE-SNCSLESKTLVYQMAVASKTEDVWVRIEDCPINPRSGRSPGTN 2672
Db
2679 TPVIDSVSEKANPNIKDKONQKQNVGNSGSPVMTVGLNRLTSPFIQVDAPQKGTET 2738
Qy
2673 SPFVIDNVDQOKE-EAAKDCHTRHSGNGVPL-LENRQKSFIKVDGLDTKGTDP 2727
Db
2739 KPGQNNPVPVSTNESPVERTPPSSSSSHSSPSGTVAAARVPFNPNPSPRKSSADST 2798
Qy
2728 KSLINQ--OBTNENTVAERTAFSSSSSHSSPSGTVAAARVPFNPNPSPRKSGENS 2784
Db
2799 SARPSQIPTPVNNNTKRDSTKDTSESSGTSQSPKSHSGSYLTVS 2843
Qy
2785 TSRRSQIPTPVNTSKRDSTKDTSESSGTSQSPKSHSGSYLTVS 2829
Db
Q8BNP7 PRELIMINARY; PRT; 1056 AA.
AC Q8BNP7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Adenomatosis polyposis coli (Fragment).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Adipose tissue;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK080907; BAC38073.1; -.
DR MGD; MGI:88039; Apc.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0008013; F:beta-catenin binding; IDA.
DR GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
DR GO; GO:0009798; P:axis specification; IMP.
DR GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
DR GO; GO:0016055; P:wnt receptor signaling pathway; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg. 4.
DR PROSITE; PS00176; ARM_REPEAT; 1.
FT NON_TER 1056 1056
SQ SEQUENCE 1056 AA; 117633 MW; 86C79FDC12C23FAB CRC64;

Query Match 35.4%; Score 5156; DB 11; Length 1056;
Best Local Similarity 94.9%; Pred. No. 3.2e-246;
Matches 1004; Conservative 23; Mismatches 29; Indels 2; Gaps 2;
Qy 1 MAAASDQLLKQVEALKMENSNLQLEDNSHLTKLETEASNKVEVLKQLOGSIEDEAM 60
Db 1 MAAASDQLLKQVEALKMENSNLQLEDNSHLTKLETEASNKVEVLKQLOGSIEDETM 60
Qy 61 ASSCQIDILERIKELNLDSSNPGVKLRKMSIRSYGSRGSSVSSGSCSPVPMGSPFR 120
Db 61 -TSGQIDLLERLKKEFLD-SNPPGVKLRKMSIRSYGSRGSSVSSGSCSPVPMGSPFR 118
Qy 121 RGVNGSRSESTGYLEBELEKERSILLADLDKEEKEDWYVYQALQNLTKRIDSPLTENFSL 180
Db 119 RTFVNGSRSESTGYLEBELEKERSILLADLDKEEKEDWYVYQALQNLTKRIDSPLTENFSL 178
Qy 181 QDLTRQLEIYEARQIRVAMEEOLGTCQDMKRAQRIARIQOIEKDILIRIOLLOSQAT 240
Db 179 QDMDTRQLEIYEARQIRVAMEEOLGTCQDMKRAQRIARIQOIEKDILIRIOLLOSQA 238
Qy 241 EADRSNKHETGSHDAERQNEQGVGEINMATSNGQGSTTTRMDHETASVLSSSSTHSA 300
Db 239 EADRSQSRHDAASHAGQHEGHGVAESNTAASSSQSPATREVDHETASVLSSSSTHSA 298
Qy 301 PRELTHLGTKEVMYVLSLLSMGLTHDKDDMSRTLLAMSSQSDSCISMROSGCLPLLIQLL 360
Db 299 PRELTHLGTKEVMYVLSLLSMGLTHDKDDMSRTLLAMSSQSDSCISMROSGCLPLLIQLL 358
Qy 361 HGNKDSVLLGNRSGKEAPARASAAALHNIHSOPDKKGRREIRVHLLEQIRAYCETC 420
Db 359 HGNKDSVLLGNRSGKEAPARASAAALHNIHSOPDKKGRREIRVHLLEQIRAYCETC 418
Qy 421 WEWQEAHEPGMDQDNKMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
Db 419 WEWQEAHEPGMDQDNKMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 478
Qy 481 VDCMYGLTNDHYSITLRRYAGNALNLTFDGVANKATLCMKGCMRALVAQKSSSEDL 540
Db 479 VDCMYGLTNDHYSITLRRYAGNALNLTFDGVANKATLCMKGCMRALVAQKSSSEDL 538
Qy 541 QQVIASVLRNLSWRADVNSKTLREVSGVKALMECALEVKEKSTLKSLSALWNLSAHT 600
Db 539 QQVIASVLRNLSWRADVNSKTLREVSGVKALMECALEVKEKSTLKSLSALWNLSAHT 598
Qy 601 ENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILNVSLSLATNEDHQILRENN 660
Db 599 ENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILNVSLSLATNEDHQILRENN 658
Qy 661 CLQTLLOHLKSHSLTIIVSNACGTLNLSARNPKDQALWDMGAVSMLKNLIHSHKMIAM 720
Db 659 CLQTLLOHLKSHSLTIIVSNACGTLNLSARNPKDQALWDMGAVSMLKNLIHSHKMIAM 718
Qy 721 GSAALRNLMANRPAKYKADANIMSPGSSPLSLHVRKOKALEAEALDAOHLSETFDNI 780
Db 719 GSAALRNLMANRPAKYKADANIMSPGSSPLSLHVRKOKALEAEALDAOHLSETFDNI 778
Qy 781 PKASHRSKQSHKQSLYGDYVFDTRHDDNDSDFNTGNMTVLSPLYNTTVLPSSSSSRGS 840
Db 779 PKASHRSKQSHKQSLYGDYVFDTRHDDNDSDFNTGNMTVLSPLYNTTVLPSSSSSRGS 838
Qy 841 LDSRSSEKRSLEERERGIGLGNYPATENPCTSSKRLQISTTAAQIAKYMEEVSAIHTS 900
Db 839 LDSRSSEKRSLEERERGIGLGNYPATENPCTSSKRLQISTTAAQIAKYMEEVSAIHTS 898
Qy 901 QEDRSSSTTELHCVTERNALRSSHAAHSHNTYNTKSENSNRTCSMPYAKLYKRSS 960
Db 899 QEDRSSASTTFEHCVAADRSAARSSASHTHSNYNTKSENSNRTCSMPYAKLYKRSS 958
Qy 961 NDSLNSVSSNDGYKRGQMKPSEIESYEDDESKEFCSYGQYFADLAHKIHSAHMDNDGE 1020
Db 959 NDSLNSVSSNDGYKRGQMKPSEIESYEDDESKEFCSYGQYFADLAHKIHSAHMDNDGE 1018

1344 SESARHKAVETPPSGAKSPKSGAOTPKSPPHYVQETPLMFESRCTSVSSILDSFESRSIAS 1403
1129 VEDA-----TPSSSSNCVQETPLVLSRCSVSSVSSGCSFESRSIAS 1168
1404 SVQSEPCSGMVSGIISPSDLPSGQNTMPPSRKTTPPPPTQATKR-----EVPKKA 1457
1169 SIPSDPCSLGSGTVPSELPSGQNTMPPSRKTTPPAPPQGPSTQFSLQWESYVXRFL 1228
1458 PTAERKESGPKQAANAAVQVQLPDATLLHATATESTPDGFCSSLSALSIDPEFIO 1517
1229 DIADRCRCQPSBELAGSVR-----FTVEKPDENFSCASLSALAHELYVQ 1276
1518 KDVELRIMPV-VQENDNGNETESQPKESNNQKEAKTIDSEK---DLLDSDDDDDIE 1573
1277 QDVELRLRPPACPERAVGG-----GCHRRRDEAASRLDGPAPAGSARSATDKELE 1327
1574 ILEECIIISAMTKSRKKGKPAQASKLPVVPVASKPSOLPVYKLLPSONRLQPOKHVSFT 1633
1328 ALRECLGAAMFARUK-----VASALVP---GRSLPVPVIMLVAPAR----- 1368
1634 PGDDMPRVYCEGTPINFSTATSILDTIESPPNELAAGEVGGAGSAGEFEKEDTIPT 1693
1369 -GDD-SGTDSEAGTVPNFSSAASLDELQPSRDKPAGPGR-----QK-----PT- 1413
1694 GRSTDEAGGKTSSVTIFELDDNKAEGDIIAECINSAMPKKGKHKPVVKIMDQVQAA 1753
1414 GRAAPARQ-----TRSHRP-----KAA 1430
1754 SASSAPKNQLDGKKKXPTSPVKIPONTY-RTRVRKNADSKNNLNAERVFSDNKDSK 1812
1431 GAKS-----TEHTRGFCNRLAGLEPLSRPOSASNRDSS 1466
1813 KONLKNKSKDFNDKLPNNEDRVRGFAFDSPHHYTPTEGTYPYCFNRNDSLSLDFDDDDV 1872
1467 CQT-----RTRGDGALQSLCLTPTTEAVYCF-----YDSDEE 1499
1873 DLS-----REKAEIRKAKENKSEAKVTSHTELTSNQOSANKOAIKQPIN---RQOP 1923
1500 PPATAPPERRASAIPLAKREKPAK-----ETPSRAAQATPLVRAQP 1544
1924 KPILOKQSTFPOSSKDIEDRGAATDKLQNFALIENTPVCFSHNSLSLSLSDIDQENNNK 1993
1545 RLJ-----VDETPPCYSLTSSASSLSLSE----- 1566
1984 NEPIKETPPDSCQBPSPQASGYAPKSFHVEDTPVCFPSRNSLSLSLSDIDEDLLQRCI 2043
1567 -----PEA---PEQPANHARGFEQGSKOD-----SSPSRAEEELLQRCI 1603
2044 SSAMP-----KKKPKPSRLKGDNEKHSRNMGGILGEDITLKDQIQRPDSE-HGLS 2093
1604 SLAMPERRTQVPSRRRKPRALRS-----IRPTEIT-----QKQEBEAVGSD 1646
2094 P-DSENFOWKAIQEGANSIVSLSHOAAAACLSRQASDSDSILSKSGISLGSPFHLT 2151
1647 PASDLDSVEWQAIQEGANSIVTLHQAAKASL--EASSESLSLSLVSGVSGAGSTLQ-- 1702
2152 PQOBEKPFPTSNKGPRLTKPEKSTLTETKI-BSESGIKGGKVKYKSLITGKVR-SNSEI 2209
1703 -----PSKLRKGRPAEAGGAWPEKRGTTSTK-----INGSPRLPNGPE 1743
2210 SQGMKQPLQANMPSISRGHTMIHIPOVRNSSSTSPVSKGPPKLPKTPASKSPSEQATTT 2269
1744 KAKGTQKMWAGESTMLRGRTVIY-----SAGPASRTQSKG--ISGPGCTTPKKTGTSGT 1795
2270 SPRGAKPSVKSELSPVARQTSIGGSKKAPSGSGSDSTPSPAPQPLSRPQSPGRNSI 2329
1796 QPETV-----TKAPSPQQRSELHRPCKISELAURHPPRSAT 1834
2330 SPGRNGISPPNKLSQLPRTSSP-----STASTKSSG-----SGKMSYTSFGRQMSQOQNTK 2380
1835 PPARLAKTFSSSSSQTSASQPL-PRRSPLATPTGGPLPGPGSLVPKSPARAL-----LAK 1890
2381 QTGLSKNASSI-----PRSESASKGLNOMNNGA-NKKVELSRMSSTK 2423

Db 1891 QHKTQKSPVKIPMORPARRVPPPLARPSPFPGSRGAGAEPTGARGSLGLVRNASAR 1950
Qy 2424 SSGSESDRSERPVLRQSTFIKEASPTLRRLKLESASPESLSPSRPASPTRSQATPV 2483
Db 1951 SSGSES--SDRSGFRQLTFIKESPG-LURRRSELSSADSTASTISOAASPRGRPALPA 2007
Qy 2484 L---SPSLPDMLSLTHSSVQAGWRKLPNLTPTIYNDGRFAKRDHDIARSHSESRLP 2540
Db 2008 VFLCSSRCDELVRSPQPLAAQKSPQAKPLAPL-----APR-----RTSESSESRLP 2055
Qy 2541 INRSGTWKBEHSHSSSLPRVSTWRTGSSSILSASSSESSEKAKSEDEKHVNSIS--GT 2598
Db 2056 V-RASGRPETVKRYASLPHISVSRKSDSAVSPTTQANATRRGSGEARPLPRVAPPGT 2114
Qy 2599 KQSKENQVSAKGTWRKIKENEF-----PTNSTSQTVSSGATNGAESKTLIYQAPAV 2651
Db 2115 -----TWRIKDEDVPHILRSTLPATLPLRVSSPEDSPAGTPQ----- 2153
Qy 2652 SKTDEWVRIEDCPINPRSGRSPT--GNTPPVIDSVSEKAMPNIKDKONQAKQNVGNG 2709
Db 2154 RKTSDAVQOTEDVATSKTNSSTSPSLESRDPP-----QAP-----A 2189
Qy 2710 SVMRTVGLNRLTSFIQVDAPQKGTETKPGONNPVSETNESPIVERTPPSSSSSK 2769
Db 2190 SGVAPQGS-----VDGP-----VLTKPPASAPFP--HEGLSAVIAGFP-----TSR 2230
Qy 2770 HSSPSGTVAARVTPFNPNPSPRKS--ADSTSRAPSIQPTPVN 2810
Db 2231 HGSPSR--AARVFPFNKVPSPMAATWASDAVEKAPVSSPAS 2271
RESULT 4
O95996 PRELIMINARY; PRF; 2303 AA.
AC O95996;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE APCL protein.
GN APCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCB1_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Koyama K., Nakagawa H., Nakamura Y.;
RT "APCL exon14";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9904063; PubMed=9823329;
RA Nakagawa H., Murata Y., Koyama K., Fujiyama A., Miyoshi Y., Monden M.,
RA Akiyama T., Nakamura Y.;
RT "Identification of a brain-specific APC homologue, APCL, and its
interaction with beta-catenin.";
RL Cancer Res. 58:5176-5181(1998).
DR EMBL; AB022529; BAA75469.1; JOINED.
DR EMBL; AB022518; BAA75469.1; JOINED.
DR EMBL; AB022519; BAA75469.1; JOINED.
DR EMBL; AB022520; BAA75469.1; JOINED.
DR EMBL; AB022521; BAA75469.1; JOINED.
DR EMBL; AB022522; BAA75469.1; JOINED.
DR EMBL; AB022523; BAA75469.1; JOINED.
DR EMBL; AB022524; BAA75469.1; JOINED.
DR EMBL; AB022525; BAA75469.1; JOINED.
DR EMBL; AB022526; BAA75469.1; JOINED.
DR EMBL; AB022527; BAA75469.1; JOINED.
DR EMBL; AB022528; BAA75469.1; JOINED.
DR EMBL; AB012162; BAA34611.1; --
HSP; Q02248; 3BCT.

DE Adenomatosis polyposis coli.
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK045053; BAC32198.1; -.
DR MGD; MGI:88039; Apc.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0008013; F:beta-catenin binding; IDA.
DR GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
DR GO; GO:0009798; P:axis specification; IMP.
DR GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
DR GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 489 AA; 55218 MW; F7C994968B01993 CRC64;
Query Match 15.3%; Score 2229; DB 11; Length 489;
Best Local Similarity 92.8%; Pred. No. 3.8e-102;
Matches 440; Conservative 12; Mismatches 20; Indels 2; Gaps 2;
Qy 1 MAASVDLLKQVEALKMENSINRQLEDSNHLTKLETSNMKEVLKQSGIDEAM 60
Db 1 MAASVDLLKQVEALKMENSINRQLEDSNHLTKLETSNMKEVLKQSGIDEAM 60
Qy 61 ASSGQIDLLERLKLNDLSNFPVGLRSLRSYSGREGSVSSRSGCSPVPMGSPFR 120
Db 61 -TSQIDLLERLKEFNLD-SNFPVGLRSLRSYSGREGSVSSRSGCSPVPMGSPFR 118
Qy 121 RGVNRSRGSTGLLELEKERSLLADLKEEKEKWYIAQLONLTKRIDSILPTENESL 180
Db 119 RTFVNSRGSTGLLELEKERSLLADLKEEKEKWYIAQLONLTKRIDSILPTENESL 178
Qy 181 QTLDTROLVEARQIRVAMEEQLGTCQDMEKQAQRRIARIQOIEKDIIRQLQSQA 240
Db 179 QTNTRQLVEARQIRVAMEEQLGTCQDMEKQAQRRIARIQOIEKDIIRQLQSQA 238
Qy 241 EAERSQMKHETGSHDAERQNEGQVGEINMATSGNGQGSTTRMDHETASVLSSTHSA 300
Db 239 EAERSQSRHDAASHEAGRQHEGHEGVAESNTAASSSQSPATRVYDHETASVLSSTHSA 298
Qy 301 PRRLTSLGTVKENVYSLLSVLGTHDKDMSRTLLAMSSQDSCISMRSQGCPLLIQLL 360
Db 299 PRRLTSLGTVKENVYSLLSVLGTHDKDMSRTLLAMSSQDSCISMRSQGCPLLIQLL 358
Qy 361 HGNDKDSVLLGNRSGSKAARASAAALHNTIHSQPDCKRGRREIRVHLLEQIRAYCETC 420
Db 359 HGNDKDSVLLGNRSGSKAARASAAALHNTIHSQPDCKRGRREIRVHLLEQIRAYCETC 418
Qy 421 WEMOEAHEPGMDQKNPMPAPVEHICPAVCVLMKLSFDEEHRHMMELGLQA 474
Db 419 WEMOEAHEPGMDQKNPMPAPVEHICPAVCVLMKLSFDEEHRHMMELGLKKA 472

RESULT 6
Q9UBZ1 PRELIMINARY; PRT; 733 AA.
AC Q9UBZ1;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

Qy 1931 STFPQSKDIPDRGAATDEKLO-NFAENTVPFNSHNSLSLSLSDIQENNNKENPEIKE 1989
Db 1546 KEAPAPSKAAP--AAPPAKTOPSLIADETPFCVLSLSSASSLS-----EPEP--- 1591
Qy 1990 TEPPDSQGEPSKPOASGYAPKSFHEVTPVCFNSNLSLSLSIDSEDDLLOECISSMPK 2049
Db 1592 SEPFAVHPGREPAVT-----KDPGPGGGRDSSPSP---RAABELLQRCISSALPR 1639
Qy 2050 KKKP-SRLKGNENKHSRNMGGILGEDLTLDLXDIQRP---DSRHG-----LSPDEN 2098
Db 1640 RRPVPSGLR-----RRKKP-----ERPAGSRERGEAAGSDRASLDLS 1683
Qy 2099 FDWKAIOEGANSIVSLHQAACLRQASSDSLSLSKSGISLGS-----PFHLPDPQ 2154
Db 1684 VEMRAIOEGANSIVTLHQAACLRQASSDSLSLSKSGISLGS-----LSPDEN 1740
Qy 2155 BEKFTSKNGRILKPKGEKSTLETKIESKGIKGGKVKYSLITGKVRNSHISQMK 2214
Db 1741 AEGEMGSARR-----EKGAASVKTGSPRSPAGPEK-----PRGT 1777
Qy 2215 OPLQANMPSISGRTHMIH-----PGVRNSSSSTSPVSKKGPPLKTPA--SKS 2260
Db 1778 QKTFGVPAVLGRITVIVPSPAPAPQPKGTGPRATPRKVP-----PCLAQAPAAK 1832
Qy 2261 PSEGTATTS--PRGAKPVKSELSPVARQTSQIGSSKAPRSRGRDSTPSRPAQPLSR 2319
Db 1833 PSPGQORSRLHRAKTSSELATLSQPPRSATPPARLAKTPSSSSQTS-----PASQPLR 1888
Qy 2320 ---PLOSGRNISIPGRNGISPPNKLISQTPSTASTKSSGSKMYSVSPGRMSQ 2376
Db 1889 KRPPVQAA-----CALPGQASVPVKTPTATLLAKHK-----TQSPVRIPFMQ 1934
Qy 2377 NLTKGTGLSKNASSIPRESASKGLNQMNGGA-NKKVELSRMSSTKSGSGESDRSERP 2435
Db 1935 RPARR-GPPLARAVP--EPGPRGAGT-EAGFGARGGLGLVVRVASALSSGES--SDRS 1989
Qy 2436 VLVROSTIKAPSTILARKLEESASFESLSPSPSPASPTRSQAQTPVLSPLDMSLS 2495
Db 1990 GFRRLTTFKE--SPGLRRRSELSAASAPQAGSPRRGR-----PALPAVFLCS 2040
Qy 2496 HSSVQ-----AGGWKRLPNLSPTIYNDGRPAKHDIARSHSESFSLPINRSQTW 2547
Db 2041 SKCELRAPROGPAPARQPPAARPS--PGERPAR-----RTTSESFSLPV-RAPAA 2091
Qy 2548 KREHKSLSLPRVSTWRTGTSSSILSASSSEKASEKHEKVNISIGTKQSKENQVS 2607
Db 2092 RPETVKRVASLPHISVARRPDGAVPAAPASADARRSSDGEPRPL-----PRVA 2140
Qy 2608 AKG-TWRKIKENEF-----SPNTSQTSSCATNGABSCTLIQMAPAVSKTEDVWVR 2661
Db 2141 AGTTWRRIRDVPHILRSTIPALPLRGSTPEDAPA-----GPPPKTSDAVVQT 2193
Qy 2662 EDCPINNPSRSGSPGNT--PPVIDSVSEKANPNIKSKNOAKONVGNVGVPMRTVGLE 2719
Db 2194 EEAAPKNTSSTPSLETRPPGAPAGGO-----LSSLGSD 2229
Qy 2720 NLTSLFIQVADPQKTEIKPQNNPVPVSETNESPIVER---TPFSSSSSKHSRSGT 2776
Db 2230 -----VDGSLA-----KAPIS-----APFVHEGLGVAVGGFPASRHGSPSR- 2266
Qy 2777 VAARVTPFNPNPSRKSADSTSA 2800
Db 2267 -SARVFPFNVPSPMVAATDSA 2289

RESULT 5
Q9BRD8 PRELIMINARY; PRT; 489 AA.
AC Q9BRD8;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

[illegible][illegible]

Db 168 ----- 167
 QY 359 LLHGNDKDSVLLGNSRGSKEARASAAALHNIHSQDDKRGREIRVLHLLLEQIRAYCE 418
 Db 168 ----- 167
 QY 419 TCWEQEAHEPGMDOKNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGIQAIAEL 478
 Db 168 ----- EL 169
 QY 479 LQVDCMYGLTNDHYSITLRYAGMALTNLTFDGVANKATILCSMKGMALVAOLKSESE 538
 Db 170 LOVDYEMHNTDPLNAL-RYAGMTLNLTFDGVANKATILCARRGMEALVAOLADSE 228
 QY 539 DLOQVIASVLRNLWRADVNSKTLREVGSKALMECALEVKVSTLKSIVLSALWNL5AH 598
 Db 229 ELHQVSSGILNLSWRADINSKLLREAGSVTALVQCVRATKESITLKSIVLSALWNL5AH 288
 QY 599 CTENKADICAVDGALEFVGLTYRSOTNTLAIIESGGILNYSLLIATNEDHQILRE 658
 Db 289 STENKAAICQVDGALGFVLSTLYTCQNSLAIIESGGILNYSLLIATNEDHQILRE 348
 QY 659 NNCLQTLQHLKSHSLTIVSNACGTLWNL5ARNPKDQBALMDGAVSMKLNLIHSHKXMI 718
 Db 349 HNCLOTLLQHLTSHSLTIVSNACGTLWNL5ARSARDQELLWDLGAVGLRLNVSHKXMI 408
 QY 719 AMGSAALRNLMANPAKYK-DANTMSGSSLP5LHVKKQKALEBAELDAQHSETFDNID 777
 Db 409 AMGSAALRNLLAHPAKHQAATAVSPGSCVPSLYVRKQALEBAELDARHQALEHLE 468
 QY 778 NLSPKASHRSKO-----RHQSLXGVDYFTNRHDDNRDNF-----NTGNMTVLS 823
 Db 469 KGPFAAEATKPLPPLRHLDLQADYASDGCDFDDDDAPSSLAATAATGEPASPAALS 528
 QY 824 PYLNTVLPSSSSSR-----GSLDSSSEKDRSLERERIGLGNYPATENPGTSGKRGLOI 880
 Db 529 LFLGSPFLOQAQARTPPRRGKEAKDTSGE-----AAVAKAKAKL 572
 QY 881 STTAQIAKMEVEGAHTSQDRSSGTEHLCHVTDERNALRBSAAHTS--NTYNT 938
 Db 573 ALAVARIDLOVEDISALHTSDSDSFLSS-----GDFGQAPREGRAQSCPCRGPEG 626
 QY 939 KSENNRTCSMPYAKLEYKRSNDSLSNVSNDGYGKQKPKSTIESYEDDESKFCYSY 998
 Db 627 REAGSRAHPLRLKAAHASLNSDLSNGSASDGYCPREHM----- 667
 QY 999 QYPADLAHKIHSANHMDNDGELDTPIYLSKYDEQLNSGRQSQSONERWAPKHIED 1058
 Db 668 -LPCPLA-----ALASREDDP----- 682
 QY 1059 EIKQEQROSRNQSTTYPVYTESTDDKHLKQFHPGQECVSPYRSRGANGSETNRVGSN 1118
 Db 683 -----RCQGPRESR----- 691
 QY 1119 HGINONVSOLCQEDDYDDKPTNYSERYSEEQHEEERPNYSIKNEKRVHDQPID 1178
 Db 692 -----LDDPPGCOAEPAREATSADARVRT---IKLSPTYQHY----- 727
 QY 1179 YSLKYATDIPSSKQSFPSKSSSGSSKSTHEWSSSENTSTPSSNAKQNLHPSSAQS 1238
 Db 728 -----PLLEGASRAEPLAGPI-----SPGARKQAWL---PDH 760
 QY 1239 RSQGPQKAATCKVSSINQETIQTVCVEDTPICFSRCSSLSLSSAEDEIGCNQTTQEADS 1298
 Db 761 LSKVPEKLAAPL-SVASKALQKLAQEGFLSLRCSLSSLSA-GRPGPSEGGLDLS 818
 QY 1299 ANTLQIAE-----IKGKIGTRSAEDPVSEVPAYVQHPRTKSSRLQSSLSSESARHAKVEF 1354
 Db 819 DSSLEGLEEAGPIAELEDTWRAPGATSLPVAIPAPR-----NNRGGLGVEDA----- 867
 QY 1355 PSGAKSPSKSGAGTPKSPPEHYVQETPLMFRCTSVSSLDSPESRSIASSVQEPSCGNV 1414
 Db 868 -----TPSSSENIVQETPLVLSRCSSVSSLSGSPESRSIASSIIFSEFCSGH 914

QY 1415 SGIISPDLPDGPQTMPPSRKTPP--PPQTAQTKREVPPKNAFTAKRESGPKQAAV 1472
 Db 915 SGIISPELDPDGPQTMPPSRKTPPLAPAQ-----PPEATQSLQWSEYVVRFLD 967
 QY 1473 NAAVQVQVLP---DADTLHFEATSTPDGSCSSLSALSILDEPTLOKDVLRIMPPVQ 1529
 Db 968 IADRCERCLPSLSDAGS-VRFTEKPDENFSCASSLSALHHEHYVQQDVELRLPSAC 1026
 QY 1530 ENDNGNETESEQKESNENQEAETIDSEKLLDDDDDDIEILEECIISAMPTKSSR 1589
 Db 1027 PERGGAGGAGLHFAHRRRGAGAGHGFSPSR-----RRGQELLELCGAAVAPARLRK 1081
 QY 1590 GKXKPAQATASKLPPVVARPQSOLPVYKLLPSQNRLOPKHVSFTPGDDMPVYCVSGTPI 1649
 Db 1082 -----VASGLVP--GRRALPVFVYMLVPAPAPQ-----EDDSCDTSASGTPV 1122
 QY 1650 NFSTATSLSLTITESPENELAGEVGRGAQGEFEKRTDITPTEGRST-----DEAOGGKT 1705
 Db 1123 NFSASLSLDETLOGPDPPOPGPAGR-----QRTGRTPSARQAMHHRKAGGAGR 1174
 QY 1706 SSVTIPELDNDKAEEDGILAEICINSAMPKSHKPRVKKIMDQVQOQASASSAPNKNQL 1765
 Db 1175 SAEQSPGAGKNRA-----GLEPLG----- 1194
 QY 1766 DGKKKPTSPVKPIQPQNTVYTRVRKNADSKNNLNAERFVSDNKNKQNLKNSKDFND 1825
 Db 1195 -----RPSAPA-----DKDSKPG----- 1209
 QY 1826 KLPNNEDVRGSPAFSPHHTPIETGTPYCFSRNDSL 1862
 Db 1210 -----RTRGDALQSLCLTTPTEBAVYCFYGNDSV 1239

RESULT 8
 P91667 PRELIMINARY; PRT; 2416 AA.
 ID P91667
 AC P91667
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ADENOMATOUS polyposis COLI.
 GN APC OR A-APC OR CG1451.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97144426; PubMed=8990193;
 RA Hayashi S., Rubinfeld B., Souza B., Polakis P., Wieschaus B.,
 Levine A.J.,
 "A Drosophila homolog of the tumor suppressor gene adenomatous
 polyposis coli down-regulates beta-catenin but its zygotic expression
 is not essential for the regulation of Armadillo.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:242-247(1997).
 DR EMBL; U77947; AAB41404.1; --
 DR PIR; T13825; T13825.
 DR FlyBase; FBgn0015589; Apc.
 DR GO; GO:0008013; F-beta-catenin binding; IDA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 5.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS0176; ARM_REPEAT; 2.
 SQ SEQUENCE 2416 AA; 261405 MW; 4DE3A10BE10E42A9 CRC64;

Query Match 12.2%; Score 1778.5; DB 5; Length 2416;
 Best Local Similarity 25.5%; Pred. No. 5.8e-79;
 Matches 721; Conservative 330; Mismatches 845; Indels 935; Gaps 101;
 QY 243 ERSNKHETGSHDAERQN-EGQGVGINVATSGNGQGTTR-----MDHETASVLSSS 295

Db 129 ELREWEHRSIDRPFQSAQOQDLDELPRNGGGSFASAGRPSRSEPSYITLSRFLDGD 188
QY 296 STHSAPR-----RLTSH-----LGTKEVWYVYLLMLGPHDMDVSRITLLAM 337
Db 189 APAPAPRLPKGAAMTTSFDERYTSVAEATLGSKEVCVYLLMLGNSNDPLEMAKFLLEL 248
QY 338 SSSQDSCISMRQSCCLPILLQHLGNKDSVLLGNRSGSKEARASAAALHNIHSQDD 397
Db 249 SGNAQSCATLRSOCNELLVQMHAPND-----QEVKCAEQALHNVHRHPDE 298
QY 398 KRGRRIRVLHLBOIRAYCETCEWQEAHEPGM--DQDNMPAPVHEQICPAVCVLMK 455
Db 299 KAGREAKVRLLDQIVDYCSFLKTLQSGGEAIDSDRHL-----AAISSLMK 349
QY 456 LSPDEEHRHAMELGGIQAETAEILOVDCMYG-LTNDHYITLRRVAGMALTNLTGDDVA 514
Db 350 VSPDEEHRHAMELGGIQAETAEILOVDCMYG-LTNDHYITLRRVAGMALTNLTGDDVA 514
QY 515 NKATLCMKGMALVAQLKSESEDLQOVIASVLRNLNLSWRADVNSKKTLEVEGSKVLM 574
Db 410 NKALLCGOKPMEALVAQLDAPDLDLQVTVASVLRNLNLSWRADSNKAVLNEICTVALLAL 469
QY 575 CALEVKKESTLKVLSALNLSAHCTENKADI CAVDGALAFVGTUTYRSQTNLAIIES 634
Db 470 AAMENRSNTLKAILLSALNLSAHCTENKAEFCVADGALAFVGMLSYEGPSKTLXIEN 529
QY 635 GGGILRVVSLIANDHROILRENNCLQTLLOHLKSHSLTI VSNACGLTNLNLNLSARNPKD 694
Db 530 AGGILRVVSHIACEYRQILQHCLAILLOLKSESLTVVSNCGTLNLSARSAD 589
QY 695 QEALWDMGAVMLKNIILHSKXKANGSAAALNLMANRPKYK----DANIMSPG-SSL 749
Db 590 QKFLWDMGAVMLRSLIHSKXKANGSAAALNLMANRPKYK----DANIMSPG-SSL 749
QY 750 PSLHVRKOKALEADACHLSETFDNDNLNLSKASHRSKORHKQSLYGVYDPTNRHDDN 809
Db 650 PTLKAKAKALQELGERHFAETCDNLD----- 677
QY 810 RSDNFTGNMTVLSPLYLNTTLPSSSSRGLSDSRSEKDRSLERBERGILGNYHPATEN 869
Db 678 -----TGG-----KLDKERASSSSRRHP----- 695
QY 870 PGTSSKGLQISTTAQIAKMEVSAIHTSQEDRSSGTTTELHCVTDERNALRSSAAH 929
Db 696 -----APLRT-----SAMLTKSESRDSVYSAKSDCAYD--HJIRASASD 734
QY 930 THSNTYNTKSENSNRCTCMFYAKLEYKRSNDLSNLSVSSNDGYGRGQMKPSIESYSD 989
Db 735 AHR-----QTEATEEQPIDYVKYS-----KVXPKITDFDLE 749
QY 990 DESKFCSYGOYPADLAHKLHSAHMDNDGELDTPINYSKLYSGDEQLNGROSPQNERW 1049
Db 750 ME-----QTEATEEQPIDYVKYS-----KVXPKITDFDLE 749
QY 1050 ARPKHIEDEIKQEQRSRQSTTPVYVTESTDDKHLKFPQHFQGCVCVPSRSGANG 1109
Db 770 -----ENAKTSTY----- 778
QY 1110 SETNRVGSNIGINONVSQSLQBEDDVEDDKPTNYSERYSEEBHEEPTNYSIKYNEE 1169
Db 779 -----QETDL--DQPTDFSLRYAE-----NQIESDLDISGPAGQ 811
QY 1170 KRVDPIDYSLKAYTDIP--SSQKQSFSSKSSGOSKTEH-----WSSSENTSTPS 1223
Db 812 KSTITP-----ATVEKSGQBEILLILDDSVKCYQETDTPVVISNAASVTLDRVA 864
QY 1224 NAKRQNLHPS--SAQSRGQPO--KAATCKVSSINQETIYCYVEDTPICFSCSSLS 1278
Db 865 KADREAEVPEVREVTSEKAPKPKLQCGSGSVYTPKPNYCEEPTGPFYSYDLS 924
QY 1279 SLSSADEIG-CNQTQEADSANTLQIAETKIGTRSAEDPVSEVPAVSQHPRTKSEL 1337

Db 925 SL-----DESGKANQAIVGTD-----ADIKPKLEKQEEQ----- 954
QY 1338 QGSSLSUSBSARHKAVERFPSPGAKSPKSGAOPPKPPEHYVQETPLMPSRCTSVSSLSDFE 1397
Db 955 -----SOPAEQVLTKPPTQANS-----ALETPLMFSRRSSMDSLVHDP 992
QY 1398 SRSLA-----SSVQSEPCSGMVSGIISPSDLDPSPGCTMPPSRSKTTPPPQTATKREV 1452
Db 993 DVDVANDCKSKVSD--FSRLASGVISPEIPDPTQSMPOS----- 1033
QY 1453 PKNKAPTAEKRESGPKQAANVAQVQLVPDADTLHFATETSTPDGFCSSSSLSALS 1512
Db 1034 PRNSVAGSQNVDSPPVIFASLOPLRSVFE--DDLGSFNVHTPAQFSTATSLNLSI- 1091
QY 1513 BPFIQDKVELRIMPVQENDNGNE-----TESEQPKESNENQKE-AEKTID 1558
Db 1092 -----VDDEKAPAVMTEDNEDELLANCINMGQKPTKPTAVKSTVVNSEVDVAETIR 1144
QY 1559 S-----EKD-----LLDD--SDDDDI----- 1572
Db 1145 SYCTEDTALLSKVPSNTNLSVISMSSTDPKDATAGQAQVYAHQSLDVSNSASDCGAS 1204
QY 1573 -EILEBECIISAMPTKSSRKGKPAQTAASKUPPVARK-PSQLPVYKLLPQNRLQPKHV 1630
Db 1205 GHLLQCCIRDMG-----KKPLGEATSDPIAMLRGNGELPGY--LPS----- 1244
QY 1631 SFTPGDDMPRVYCVGTPINFSTATSLDLTIES-----PPNELAAGGVGRGGAQSGBEF 1686
Db 1245 -----ADENMK-FLVEDSPCFVSVGSLNLTGSSLVGPAVOLKETB-PSADQNPENK 1298
QY 1687 RDTPTGRTDEAOGGKTSSVTIPDELDDNKAEBGDIABE----- 1727
Db 1299 SLANRSKRRPPHWQDDSLSLSDSEDDTNLLSQAIAAGCNRPKSLGFSNGKRSSL 1358
QY 1728 -----INSAMPKGSHPKPRVKKINDVQOASASSASAPKQO-----LDG-KK 1769
Db 1359 SSSQPIALNAATSASSINSAMTVKSKQOQESVSDSDSDNDNOSKSLFELCILKGMKT 1418
QY 1770 KKP-----TSPVKPIPTQNTYRT-----RVRKNADSKNNLNAER 1803
Db 1419 KEPGARAQMQEQPIVGSVSSVQSNPSLKQFDSLVPQLPSSQVQKQRHHHHHRRER 1478
QY 1804 VFSNDKSK-KQNLKN--SKDFNDKLPNNEDVRGSAFDSPHHYTPIEGTVCYCSRND 1860
Db 1479 ---BRKEKLLQECINTGISKKIN-AVPRNV--LATSAAALEPCH--PMAAT-----T 1523
QY 1861 SLSLDFDDDDVDLSREKAEIRKAKENKSEAKVTSHTELTSTNQOSANKTQA-IAQPI- 1918
Db 1524 SASALSTAAPDVE--QKAH--ATSNPQOQSSHTPSSHILPNPDAIATVDTARSPAA 1577
QY 1919 -NRQOPKPILOKQSTFQSSKDPDRGAATDEKLONFAIEN-----TPVCF----- 1963
Db 1578 PNOGNGNASQNGLET-ATGSKDLSDSEDRSSDESNQSFIMETMYRLDSALNETCISGASEK 1636
QY 1964 -----SHNSLSLSLSDIDQENNNKENEPKETEPPDS 1995
Db 1637 HKDPDLMLKSVRLTMEFVTSABQLRSSSHNASS-----NSHKQNSNNNTWNESTCPND 1691
QY 1996 QGEPSPKQASGAPKSPHVEDTVPVCFSRNSSLSLSID---SEDDLLQECISSAMPKKK 2052
Db 1692 VSFPSVQTA-----PV-----LASLSDIDEATEARSLHLEIETPTNEQ 1732
QY 2053 PSRLKGNK-----HSPRNMGILGEDLTLDLQIOPDSEHGLSPDSENFDMKAIQEGA 2108
Db 1733 PESLEGFTILVNGHADSYSGSGGLNFQJ---GGQVQAGVLEFORLLFNCTASIMT 1789
QY 2109 NSIVSSLHQAAAAACLSRQASSDSD-----SILSL-----KSGI-SLGSPPF--HLTP 2152
Db 1790 NSTMIAFEARALAEALQPAATDDDTTMTFSLNSLDLDIRPPSGWESLNSQDHSQP 1849
QY 2153 DQBEKPTSKNGPRILKPGEKSTLETKIBESKGIKGGKVKYSLITGKVRNSISQ 2212
Db 1850 SSLRQAMPS--KSPRFARKMPFANLVARALGHUAG-----SAESVNSCNDLDN 1897

Db 530 AGGILRVSHIAVCEPYRQILRHQNCIAILLQOLKSESLTVVNSCGTILNLSARSABD 589
Qy 695 QEALWDGAVMMLKNLTHSKHMIAMGSAALANLWANRPAYK---DANINSPG-SSL 749
Db 590 QKFLWDNGAVMFLRSLTHSKHMIAMGSAALANLWANRPAYK---DANINSPG-SSL 749
Qy 750 PSLHVRKOKALEABDLAQHLSSETFDNIDNLSPKASHRSKORHKQSLYGDYVFDNHDN 809
Db 650 PTLERAKALQELGERHTAETCDNLD----- 677
Qy 810 RSDNFNTGNMVLSPYLNTTFLPSSSSRGLSDSRSEKORSERLBERGIGLGNYPATEN 869
Db 678 -----TGG----- 695
Qy 870 PGTSSKEGLQISTAAQIAKMBEVSIAHTSQEDRSSTTELHCYTDERNALRRSSAAH 929
Db 696 -----APLTR-----SAMLTKSRSRDSVYSAKSDCAYD--HLIRASASD 734
Qy 930 THSNYNTFKSENSNRTPCSMPYAKLEYKRSNDLSLNSVSDNGYGRGQMKPSIESYSED 989
Db 735 AHR-----KVKPKITDFDLE 749
Qy 990 DESKFCYQVADLAHAKIHSANMMDNGELDTPNYSKYSDEQLNGRQSPSONERW 1049
Db 750 ME-----QTEATEBPIDYSVKYS----- 769
Qy 1050 ARPKHIEDEIKQESRQSRNOSTTYPVYTESTDDKHLKFPQHFQOQECVSPYRSRGANG 1109
Db 770 -----ENATKSTY----- 778
Qy 1110 SETNRVGNHGINQWQSICQEDDEYDDKPTNYSERYSEEBEERPTNYSIKYNEE 1169
Db 779 -----QETDL--DQPTDFSLRYAE-----NQIESDLDISGPAQGO 811
Qy 1170 KRHVDPIDYSLKYATDIP--SSQKQSPFSKSSGQSSKTEH-----MSSSENTTSPSS 1223
Db 812 KSTITPP-----AETVPKSGEQEILLIDDSVKYCYQEDTPYVISAASVTDLRVAA 864
Qy 1224 NAKRQNLHPS--SAQSRGQPO---KAATCKVSSINQETIOTYCYVEDTPICFSCSSLS 1278
Db 865 KADAEABEVKPEVRBVTSEKAPKPLKLSQSGSVTPKEPKINPYCEGTGPGVPSRVDLS 924
Qy 1279 SLSSAEDEIG-CNQTQOEAQANTLQIAETKIGKIGTSAEDPVSFPAVSQHPRTKSSSL 1337
Db 925 SL-----DESKANQAIQVTD-----ADIKPLEKQESOE----- 954
Qy 1338 QGSSLSSESARHKAPEPPGAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVSSLSDFE 1397
Db 955 -----SOPAEQVLTKPTQANS-----ALETPLMFSRSSMDSLVHDP 992
Qy 1398 SRSTA-----SSVOSEPCSGMWGIIISPDLPSPGCTMPPPSKSTPPPPPTQATKREV 1452
Db 993 DVDVANDCKSSVSD--FSRLASGVISPSBIPDSPTQSMFQS----- 1033
Qy 1453 POKAPTAERKESGPKQAAVAVORVOVLPDADTLHPATESTPDGFCSSLSLSLSL- 1511
Db 1034 PRNSVAGSQNVDSPPVVIPLASLOPLRSVFE--DDLSSFNVEHTPAQFSTATSLNSLSIV 1092
Qy 1512 -DEFFIOQVELRIMPPVQNDNGNE-----TESEQPKESNENQKE-AEKT 1556
Db 1093 DDE-----KAPASVABEEDNEDELLANCINMGQMRKFTPEAVKSTVYNSEVDVAEET 1143
Qy 1557 IDS-----EXD-----LLOD--SDDDDI-- 1572
Db 1144 IRSVCTEDTALLSKVFNNTNLVSI SMSSTDPKDATAGQOMTAHQLSDDVSSNASDCGG 1203
Qy 1573 ---EILEECIISAMPTKSRKPKAQATASKLPPPVARK-PSOLPYVLLPSQNRLOPOK 1628
Db 1204 ASGHLLQOQIRDM-----KKPLGEATSDPIAMLRGNGNELPGY--LPS----- 1245
Qy 1629 HVSTPPDDMPRYVCVGTPIINSTATSLSLDLTIES-----PPNELAAGVGVRGAQSGEF 1684
Db 1246 -----ADEMNK-FLVEDSPCNFSVSGLSNLTVGSLVGPVQLKETE--PSSADQNPEM 1297

Qy 1685 EKRDITPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAEC----- 1727
Db 1298 KAKPKQOQVRRPHWQDDSLSSLSIDSEDDTNLLSQAIAAGCNRPKSNLGFSSNGKRSS 1357
Qy 1728 -----INSAMPKKGSHKPFVRVKIMDOVQOQASASSGAPKNQ-----LDG-- 1767
Db 1358 SLSSSQPIAINAATASASSLNSAMTVKRSQOQESVSSVDSDNDNQSKSLFELCILKGY 1417
Qy 1768 KKKXP-----TSPVKPIPONTYRT-----RVKKNADSKNNLNA 1801
Db 1418 KTKPGARAQOQOQPIVIGSSSVQSNPSLKQDFLPVQLFSSGQVQKQRQHRRHHHRRER 1477
Qy 1802 ERVFSNDKSK-KNLLKN--SKDFNDKLPNNEDRVGSPAFDPPHVTPIEGTPYCFSR 1858
Db 1478 ER-----ERDEKLLQECINTGISKIN-AVPKV--LATSAALEPCH--PMAAT----- 1523
Qy 1859 NDSLSLDDDDVDLGRKAEALRAKENKESAEKVTSHTELTNSQOSANKTOIAKOPI 1918
Db 1524 -TSASALSTAAPDV-----EQKAHATSNPQ--QSSSTHPSSHILENPI 1563
Qy 1919 -----NRGQPKPILOKOSTPQSSKDI-PDRGAATDEKLONFAEN----- 1958
Db 1564 DAIAVTDTVRSAPAPQNGNASQNGLET-ATGSKOLDSEDRSDSNOSFIMETWRL 1622
Qy 1959 -----TPVCF-----SHNSSLSSLSLSDIOENN 1980
Db 1623 DSALNETCISGASEKHKDPDLMLKSVRLTMEFVTSAEQLRSSSHNHSS-----NSHN 1677
Qy 1981 KNEPIKETEPPPOSQEPKQASGVAPKSFHVEDTPVCFSRNSSLSSLSID---SEDD 2037
Db 1678 NSSNNTWNEPCPDNDVSPFSQTA-----PV-----LASLSDEBATEAR 1718
Qy 2038 LLOECISSAMKPKKPKRLKGDNEK---HSPRNMGGILQEDTLTLKDIQRDPSEGLS 2093
Db 1719 SLHELIEITPNEQOPSELEGETTLVNGHADSVSGSGGLNFOL---GGQVQNAVRL 1775
Qy 2094 PDSNFDKWAIOEGANSIVSSLHQAALCLSRQASDSD---SLSL-----KS 2140
Db 1776 FQRLLFNGTSASIMTNTMTAFEARALAEALLOPAATDDTTMTFSLNLDLNRTPPS 1835
Qy 2141 GI-SILGSPF--HLTPDQEEKPFTSNKGPRIILKCEKSTLETKIYESKGIKGGKVKYS 2197
Db 1836 GWESLNSCYQHSQFSSLRQAMPS-KSPRPARKFPANILVARRALCHLAG----- 1884
Qy 2198 LITQKVRNSNEISQOMKOP-----LOANMPSISGRMTIHIPGVNRSSSSTSPSVKKGPP 2252
Db 1885 -SASVNSSCNLLDNKPPSLMDELDSMISVDSIQSEVADGEQDCSMATTISVSNYETA 1943
Qy 2253 L-----KTPASKSPSGOTATTSPPGAKSPVSKSEL----- 2282
Db 1944 ACDDQMTVLQSCDEDEDDATNDYSSAESTPKHGSTPPNRRSLTTPDKRRLTKDRFPT 2003
Qy 2283 -----SPVARQTSQI-----GGSKAPSRGSRDSTPSRPAQPLSRPI----- 2321
Db 2004 YTIATSCMEAPEANETLQIEVEAAVPVATPSPRANGRRGSAERYKTOLIECPALAIQ 2063
Qy 2322 -----KSP--GRNSISPGR 2333
Db 2064 PQDDCPSEQLSSIRAMWQOFTFITDINIGHSQBTCTSDHPEDAGSEPCDQNSETESC 2123
Qy 2334 NG-----ISPPNKLSQL-----PRTSSPSTASTKSGSGKMSYTSRGMQSNLTKQT 2382
Db 2124 DQOEPDQLPPPSIVDLRTSVVKTLEPATAVKLVRGRKKPAVSVYMQSQRN----- 2178
Qy 2383 GLSKNASSIPRESASKGLNQMNNGANKVELSRMSSTKS-----SGSESRSRSPV- 2436
Db 2179 -----SNNAPSKKKTUSPTIAKRSVLPGSGVRLPAKKKPTP 2216
Qy 2437 -----LVQSTFIKEAPSPTLRRKLEBSAPESLSPSRPASPTRSQAQTPVLSFSL 2488
Db 2217 PPEPAPALERQGTIFVKD-----EPINSNVQPVVE-TK 2249

```

QY 2489 PMSLSTHSSVAGGWRKLPNLSPTIEYNDGRPAKRHDIAHSHSESPRLPINRSGTWK 2548
Db 2250 FAQTSPTHRA-----SKLPTKGTGTA-----SGSFSK-----AGSPKRIPLAPA---R 2289
QY 2549 REHSHSSSLPRVTRWRTGSSSILSASSESEKAKSEDEKHVNSISGTQSKENQVSA 2608
Db 2290 RMTFORANTSLRAAGKSHAASRVVSGVSTTPPSRNSNLNGSSAAAAAAXINHAQS 2349
QY 2609 K--GTWRKIKENEFSPNLS---TSQTVSSGATNGAESTLIYQVAPAVSKTEDVWVRIED 2663
Db 2350 RIANWKVDEAKTKQSSNLRKTQKSSNMLNANGTKPTLLR-----SSTFD----- 2397
QY 2664 CPINNPRSGRPTGNTTPEVIDSVSEK 2689
Db 2398 -----NTPSTAGGVKSK 2409

RESULT 10
QY Q9UEM8 PRELIMINARY; PRT; 1685 AA.
AC Q9UEM8,
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE APC2 protein (Fragment).
GN APC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA Van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor."
RL Curr. Biol. 9:105-108(1999).
DR EMBL; AJ311187; CA010317.1; -.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR Pfam; PF00514; Armadillo_seg. 2.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
FT NON TER 1
SQ SEQUENCE 1685 AA; 175540 MW; B9B81923F3912F77 CRC64;

Query Match 11.7%; Score 1705; DB 4; Length 1685;
Best Local Similarity 28.4%; Pred. No. 1.5e-75;
Matches 636; Conservative 271; Mismatches 672; Indels 660; Gaps 85;

QY 654 QILRENNCLQTLQHLKSHSLTIVSNAGCTLWNLARSNPKDQELWDMGAVSMLKNLIHS 713
Db 1 QVLRDHNCLQTLQHLTSHSLTIVSNAGCTLWNLARSARQELLWDI;GAVGMLRLNVHS 60
QY 714 KHKMIAMGSAALRNIMANRPAYK-DANIKSPGSLPSLHVKKOKALEABLDQHLSET 772
Db 61 KHKMIAMGSAALRNLLAHPAKHQAAATAVSPGCVSPSLYVRKQRALEABLDARHLAQ 120
QY 773 PNIDNLSFKASHRSKQ-----RHQSLYGDYVDTNRHDDNRSDN-----FNTG---- 817
Db 121 LEHLEKQGPAAEATKPLPLRLHLDGLAQDYASDSCFDDDDAPSSLAATAATGPAS 180
QY 818 -----NMTVLPYLYNTTVLPSSSSRGSLDSRSEKDRSLERERIGLGNYPATENPOTS 873
Db 181 PAALSILFLGSPPLQGLARTPTPRG--GKEAKDTSGE-----AAVA 222
QY 874 SRGLQISTTAQIAKWEVSAITSOEDRSSGSGTTELHCVTDERNALRBSAAHTHS- 932
Db 223 AKAKIALAVARIDOLVEDISALHTSSDSSFSS-----GDPQGEAPREGRAQSCSP 276

```

```

QY 933 -NTYNTKSENSNRTCSMPYAKLEYKESNDNSVSSNDGYCKRGQMKPSIESYSEDD 991
Db 277 CRGPEGRREAGRAHPLRLKAAHASLNSDSGSDGYCPREHM----- 324
QY 992 SKFCSYGQVPADLAHKIHSANHMDDNGELDTPINYSLKYSDEQLNSGRQSPSQNERWAR 1051
Db 325 -----LPCPLA-----ALASRREDP----- 339
QY 1052 PKHIIEDIKQSEQRORNSQSTIYVYVTESTDDXHLKFQPHFGQOQECVSPYSRGANGSE 1111
Db 340 -----RCGQPRPS----- 348
QY 1112 TNRVGSNHNQINQVQSLOQEDDYEDDKPTNYSERYSEBEHEEERPTNYSIKYNEER 1171
Db 349 -----LDLDLFCQAEPPAREATSADARVTR--IKLSPTIYQ 382
QY 1172 HYDQPIDYSLKYATDIPSSQKQSFSSKSSQSSQSSKTEHMSSESSENTSTPSSNAKQNL 1231
Db 383 HV-----PLLEGASRAGAEPFLAGGI-----SPGARQAWL 413
QY 1232 HPSSAQSRSGQPOKAATCKVSSINQETIOTYCVEDTPICFSRCSSLSLSSAABDEIGCNQ 1291
Db 414 ---PADHLSKVPEKLAAPL-SVASKALQKLAQEGFLSLSRCSSLSLSSA-GRGPSE 468
QY 1292 TQOADSANTLQIAEIKGIGKIGTRSAB-DPVSEVPVAVSQHPRTKSKRLQSSLSSESARHK 1350
Db 469 GGLDLDSDSSLEGL---EAGPSEABLDSTWRAPGATSLP----- 505
QY 1351 AVEFPSCAKSPSKS---GACTPKSPPEHYVOETPLMFBSRCTSVSSLDSDSEBSRSIASSVQS 1407
Db 506 -VAIPARRNRGRGLGVGDATPSSSENIVQETPLVLSRCSSVSSLSGSPESPASIIPS 564
QY 1408 EPCSGMVSGIISPDLPDPSGQTMPPSRSKTPP--PPQTAQTKREVVPKNAKTAERES 1465
Db 565 EPCSGGSGTISPSLPSDPSGQTMPPSRSKTPPLAPAQ-----PPEATQFSLOWES 617
QY 1466 GPKQAAVNAARVQVLP---DADTLLHATETSTPDGFSRSSLSLSLDEPIQKDEL 1522
Db 618 YVRFIDDIADCRERCLPSLELDAGS-VRFTEVKEDNFENSCASSALAHHEHVQDDVEL 676
QY 1523 RIMPPVQNDNGNETSEQPKESNENQEAETIDSEKLLDDDDDDIIEECIIISA 1582
Db 677 RLIPSCPERGGGAGGAGLHFAHRRREBGPATGSRPR---GAADQELLELRECLGAA 732
QY 1583 MPTKSRKGGKPAQATASKLPPPVARPSQLPVYKLLPSQNLQPKHVSTPDGDDMPRVY 1642
Db 733 VPARLRK-----VASALVP--GRRALPVVTVMLVPAPAQ-----EDDSCDT 773
QY 1643 CVSGTPINRSTATSLDLTIESPPNELAAGEVGRGAQSGEPEKRTIPTEGRSTDEAOG 1702
Db 774 SAGTVPNFSASAASDELQGGPRQPGPAGR-----QRPT-----GRPTBARQA 820
QY 1703 GKTSSVTIPELDDNKABEGDILAEACINSAMPKGSKHPKPRVKIMDQVQOASASSAPNK 1762
Db 821 -----MGRHK-----AGGAGRSAEQ 836
QY 1763 NQLDGGK-----KKTPSPVKPIQNTYEYTRVKNADSKNNLNAERVPFSDNDSKK 1813
Db 837 SRGAGKNRAGLEPLGRPPSAPA-----DKDGSK 865
QY 1814 QNLKNSKDFNDKLPNNEDRVGSAFDPSPHYTPTEGTPYCFPSRNDLSLSDFFDDDDVD 1873
Db 866 PG-----RTRGDGALQSLCLTPTTEAVYCFYGNDS----- 896
QY 1874 LSREKAELRVAKENKSEAKVTSHTLTNSQOAKNTQAIKQPINRGQPKPILOKQSTF 1933
Db 897 -----DEEPFAAAPTPTH-----RRTSAIPR-AFTREPR-----QQRKEA 930
QY 1934 PQSSKIDIPRGAATDEKIQ-NPAIENTPVCFSHNSLSLSDIDQENNNKNEPIKETEP 1992
Db 931 PAFSKAAP--AAPPPARTQPSLIADETPPCYSLSSASSLS-----EPFP---SEP 976

```


RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacile J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svarskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RY Yu X., Bienz M.;
RT "A new *Drosophila* APC homolog concentrated in apical adhesion zones of
RT epithelial cells,"
RL Nat. Cell Biol. 0:0-0(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor,"
RL Curr. Biol. 9:105-108(1999).
DR EMBL; AF03746; AAF6249.1; -
DR EMBL; AF113913; AAD40227.1; -
DR EMBL; AF091430; AAD20985.2; -
DR FlyBase; FBgn0026598; Apc2;
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 6.
DR PROSITE; PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116702 MW; 20C8F5F6121888F8 CRC64;

Query Match 8.4%; Score 1221.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 6.1e-52;
Matches 429; Conservative 177; Mismatches 380; Indels 647; Gaps 50;

QY 329 DMSRTLLAWSSQDSCINRQSGCPLLIQLLHNDKDSVLGNSRGSKARARASALH 388
DB 12 ELTRNFELESLRNPETCTALRSSDCIQLLVQILHANDE-----GLSTAKKYASQALH 62

QY 389 NIIHSQDDKGRREIRVILHLEQIRAYCETC-WEMQEAHEP-GWQDQKNPAPVEHQI 446
DB 63 NIVHNNPEKEQREVKMLRLDLDQILDYCNFHTQLQSGGAEIADDERHFL----- 114

QY 447 CPAVCLMLKLSFDEEHRHAMNELGLGLOIAELLOVDCEMYG-LITNDHYSITLRRYAGNAL 505
DB 115 -AAMKLLMKASFDEEHRQTMCELGALKAPNLVHLDAHVHGAAGREQCNALRSYGLMAL 173

QY 506 TNLITGDD--VANKATLCSMKGCKRVALVAQLKSESDLOQVIALSVRLNSWRADVNSKTL 563
DB 174 TNLITGDDNVHNSXYLCGQRQFMEVVIQLNTPADELQVLQVLAGVLRNLSWRADHMKITIF 233

QY 564 REVGSVKALMECALEVKESTLKSVLISALWNLNSAHCSTENKADICAVDGAFLVGLTYR 623
DB 234 NELGTVTSLARAAMQNKNENTLKAILSALWNLNSAHCSTNKAEFCVADGALFLVGLMSYE 293

QY 624 SQNTLAIIESGGILRNVSIIATNEDHROILRENNCLQTLQHLKSHLTIIVSNAGCT 683
DB 294 GPSKTLKIENAGGILRNVSIIAVCEPYRQLIRYNCLAILLQKLSLTSWTWNSCGT 353

QY 684 LWNLSARNPKDQALWDMGAVSMLKNLTHSKHKWTAMGSAALRNLMANRPKYKDANIM 743
DB 354 LWNLSARCPEDQOYLIDHNAIPLLRALISSKNMIAEGSASALKNLNVPFRATQELMPN-- 411

QY 744 SPGSSLPFLHVRKOKALEAELDAOHLSETFDNIDNLSPKASHRSKORHKQSLYGVDFPT 803
DB 412 GDGGSLLP----- 418

QY 804 NRHDDNRSNFNTGNMTVLSPYLNTVLPSSSSRGSLDSSRSEKDRSLERERIGLGN 863
DB 419 -----LDKBAHG----- 426

QY 864 HPATENPGTSSKRGLOISTTAAQIAKVMBEVSAIHTSQEDRSSGSTTELHCVTDERNALR 923
DB 427 -----GTLPR----- 432

QY 924 RSSAAHTNTYNTKSNRNTCSMPYAKLYKVSNDLSNVSNDGYGKRGQMKPSI 983
DB 433 -----FSSLSSNPTGSLKKVRPS-----TVSTTGLFNKRKRSRESI 470

QY 984 ESYSEDESKEFCSGYGOYPADLAHKIHSANHMDNDGELDTPINYSKYDEQLNSGRQSP 1043
DB 471 YSKSD-----STKYST----- 482

QY 1044 SQNERWARKHIEDEIKOSEQRNQSTTVYVYTESTDDKHLKQPHGQOECVSPYR 1103
DB 483 -----K 483

QY 1104 SRGANGSETRVGSNHGINQVNSQLCQEDDDKPTNYSERYSEEQHEEERPTNYS 1163
DB 484 SEGA-----KNPFEIVTPT-----EQPIDVS 505

QY 1164 IKYNEEKRH-----VDQPIDYSLKYATDIPSSQKQSFSSKSSGSSGSSKTEHMSSS 1215
DB 506 MKYWEHKPNSSKTFEIDLQDPTDFSARY-----KERRSAQTPAKELKSETN 551

QY 1216 ENTSTSSNAKRONQLHPSSAQSR-----SQPQKAATCKVSSINQEIQYCYVE 1265
DB 552 EIRS-----KELQLTQSSATELRNSPGLVAVSAKQKIAT-ETETETAERPINYCEE 603

QY 1266 DTPICTSRCSLSLSLSAEDEIGCQNTQBADSANTLQIAEIKGKIGTRSAEDPVEVPA 1325
DB 604 GTFGSFRPDSLSL-----TEKPE 623

QY 1326 VSGHPTKSSRLGSSLSSESARHKAVEPPSGAKSPSKGAQTPKSPPEHYVQETPLMFS 1385
DB 624 KCMPPKTPTKT-----AVLFVHVDGNTTPQIDS--ALETFLMFS 660

QY 1386 RCTSVSLDSFERSIA-----SSVQSEPCSGMVSGIISPSDLPDPSQOTMP--PSRSKTP 1439
DB 661 RRSMDSLVG-DDETVACEDNGSVISE-YSRMQSGVISPSELPSDPTQSPQSPRRDR-- 716

QY 1440 PPPPQTQATKREVPKNKAPTAERESGPKQOAVNAAVQVRVQLPDADTLHLHFAESTPDG 1499
DB 717 --KVSTQNNLDTPEQKPSVFE-----DKLNRHFVHTPAA 750

QY 1500 FSCSSLSLSLDEPFTQKQVLELRIMPPVOENDNGNETESEQPKESNEQF-KEAEKTI 1558
DB 751 FSCATLSLSLSM-----MDDSNANAIQRGRNDNGNDGAPRSYCTED 793

QY 1559 SEKDLDDSDDDIIELECIIISAMPTKSSRKQKPAQTASKLPPVPA-----RKPSQ 1611
DB 794 TTAVLKAPNSDLSIL-----SIPNDLN-----ANEAPVPAPEADVTGMDTRMPAE 841

QY 1612 LPVYKLLPSQNRLOPQKHVFTPDGDMRVYCVGEPINPSTATSLDITIEFPNLA 1671
DB 842 DAISKMECGGNALP-----SYLPVSDMSKYVEDSECTFSVIGLSHLTV----- 887

QY 1672 GEGVRGAQSGEPKRDITPTTEGRSTDEAQGKTSSVTIPELDNKK-AEEDGDLAEICNS 1730
DB 888 -----GSAKAGPULK---LPM--RTAEAAQ-----PKLPPRSVAVQGD-----AEP 924

QY 1731 AMPKXGSHKPRVKKIMDQ-----VQOASASSAPKNQDLGKKKKKTPSPVKPIPQNTYR 1786

```

Db 925 RLPPKSDSSJSMDDDDCNLLSQAIAAGSC-----RPOFSGA----- 964
Qy 1787 TRVEKNADSKNNNAEAVFSDNCKSKQNLKNNKSFNDKLPNNEDRVGSPAFDSPHHY 1846
Db 965 -----STSSSLANASTSTLCRENGSKKQ-----VEHGDK-PN----- 996
Qy 1847 TPIEGTFCFRNDLSLSLDDDDVDLSREK-----AELRKAKENKESA----- 1892
Db 997 -----YSSDDSL-----DDDDDDARSKSLFEQCILSGMHKSNDALESEGEPPGOR 1041
Qy 1893 -KVTSHTELSNQ 1904
Db 1042 QETISARDRFVSNQ 1054

RESULT 13
Q961B0 PRELIMINARY; PRT; 1067 AA.
AC Q961B0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE L024920P.
GN APC2 OR CG6193.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051719; AA93143.1; -.
DR FlyBase; FBgn0026598; Apc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 7.
DR DR PROSITE; PS00176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116678 MW; 03C4119AEF19D198 CRC64;

Query Match 8.4%; Score 1220.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 6.9e-52;
Matches 429; Conservative 177; Mismatches 380; Indels 647; Gaps 50;

Qy 329 DMSRTLLAMSSQDSICSMRQSGCLPLLQLLGHNDKDSVLLCNSRGSKEARASAAHLH 388
Db 12 ELTRNFLELSRNETCTALRSSDCIQLLVQLHANDE-----GLSTAKKYASQALH 62
Qy 389 NIITHSPDDKGRRETRVLHLEQIRAYCETC-WEHQEAHEP-GMDQDKNMPAIVEHQI 446
Db 63 NIVHNPPEKERQREVMKRLDQLDYCNFLHTQLQSGGEAIADEDRHPL----- 114
Qy 447 CPACVLMKLSFDEERHMANELGGLQIAELLOVDCEMYG-LITNDHYSITLRRYAGMAL 505
Db 115 -AAKLLMKASFDEERHQTMCGLKALPAIPLVHLHDHVAHGPAAGNEQCNALRSTGLMAL 173
Qy 506 TNLTFGD--VANKATLCMSKGNRALVAQKSESEDLOQVIASVLRNLNLSWRADVNSKTL 563
Db 174 TNLTFGDENVHNSYLCGQRFQEVVIAQLNTPADELLQVLAVLRNLNLSWRADKHKMTIF 233
Qy 564 REVGSVKALMECALEVKKESTLKSLSLWNLNLSAHCENKADICAVDGLALAEVGLTYR 623
Db 234 NELGTVTSLARAQNKENTLKALUSALWNLNLSAHCSTNAECFAVDGALAEVGLMSTY 293
Qy 624 SQNTNLAIIESGGGILRNVSLSIATNEDHRQILRENNCLQTLLOHLKSHSLTIVSNACGT 683

```

```

Db 294 GPSKTLKIIENAGGILRNVSISHIACEPYRQILRRYNCCLAILLQOLKSESITVVSNSCGT 353
Qy 684 LWNLSARNPKDQEAALWDMGAVSMKLNLIHSHKMIAMGSAALRNLANRPAKYKDANIM 743
Db 354 LWNLSARCPEDQOYLIDHNAIPLRALISSKNSMIAGSASALXNLNFRATQBLMFX-- 411
Qy 744 SPGSLPSLHVRKQKALEAEADLAQHLSETFDNIDNLSFKASHRSKQHKOSLYGDIYFDT 803
Db 412 GDGGSLLP----- 418
Qy 804 NRHDDNRSDNFTGNMTVLSPLYLNTTVLPSSSSSRGSLDSRSSEKDRSLERERIGILGNY 863
Db 419 -----LQKBAHG----- 426
Qy 864 HPATENPGTSSKRGLOIQTAAQIAKYMEEYSAIHTSQEDRSSTTTELHCVTDERNALR 923
Db 427 -----GTLPRR----- 432
Qy 924 RSSAAHTSHNTYNTFKSENSNRTCSMPYAKLEYKRSNDSNLSNVSNDGKYGKQMKPDI 983
Db 433 -----FSSLRUSNPTGSLKKVRPS-----TVSTTGFLNRKCESESI 470
Qy 984 BSYSEDDDESKFCSYGQYPADLAHKSANHMDDNDGELDTPINYSKYSDQLNSGRQSP 1043
Db 471 YSGKSD-----STKYST----- 482
Qy 1044 SQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHLKQPHFGQOECVSPYR 1103
Db 483 -----K 483
Qy 1104 SRGANGSETNRVSGHGINQVNSQLCEDDYEDDKPTNYSERVSEBEQHEEBERTPNYS 1163
Db 484 SEGA-----KNPFEIVTPT-----BEQPIDYS 505
Qy 1164 IKYNEEKRH-----VDQPIDYSLKAYATDIPSSQKQSFSSKSSGSGQSKTEHSSSS 1215
Db 506 MYMEHKPNSSKTFEIDLDQTFDSARY-----KERRSAQTAQPELSEKN 551
Qy 1216 ENTSTPSNAKRONLHPSSAQSR-----SGOPOKAACTKVSSINQETIYQYCE 1265
Db 552 EIRS-----KELQLTSSSATELRNSPGLNVAASAQKQIAT-ETETETAERPINYCEE 603
Qy 1266 DTPICFRCSSLSLSLSSAEDEIGCNQTTQBADSNLTQIAEIKGKIGTRSAEDFVSEVPA 1325
Db 604 GTPGFSRFDLSNL-----TEKPE 623
Qy 1326 VSQHPRTKSSRLQSSLSSESARHKAVERFPGSAKSPSKGAQTPKPPPEHYVQETPLMFS 1385
Db 624 KCMPPKTPKT-----AVLPVHVDGNTFQNIIDS--ALETFLMFS 660
Qy 1386 RCTSVSLDSFEGERSIA-----SSVQSEPCSCMVSGIISPSDLDPSPQOTMP--PSRSKTP 1439
Db 661 RRSMDSLVG-DEETVACEDNGSVISE-YSRQSGVISPSELDPSPQSMQSPRRDR-- 716
Qy 1440 PPPQTAQTKREVPKNKAPTAEKRESGPKQAANAARVQVLPDADTLHFAETESTPDG 1499
Db 717 ---KVSTQNKLDTPEQKPFSTVFE-----DKLNRHFVEHTPAA 750
Qy 1500 FSCSSLSLSLSDLEPFTQKDVLRIMPVVOENDNGNETSEQPKENQOE-KEAEKTIID 1558
Db 751 FSCATSLSNLSM-----MDDSNANAIQRQNDINGNGDAPRSCTED 793
Qy 1559 SEKDLLDDSDDDIEILEECIIISAMPTKSRKKGKPAQTAASKLPPVPA-----RKPSQ 1611
Db 794 TTAVLKRAPNSDLSIL-----SIPNDLN-----ANEAPVPAADVGTGMDTRMPAE 841
Qy 1612 LPVYKLLPSNRLQPKQKHSFTPGDMPRVYCVGEGTPINFSTATISDLTIESPPNELAA 1671
Db 842 DATSKMCGGNALP-----SYLPVDEMCKYYVEDSCTFSVLSGLSLTV----- 887
Qy 1672 GEGVRGGAQSGEPEKRDITPTEGRSTDEAQQGKTSSTVITPELDDNK-ABEGDILAECINS 1730

```

Db 888 -----GSAKAGVFLK-----LPM--RTAEAAQA-----PKLPPRRSAVQGD-----AEP 924
QY 1731 AMPGKSHKPRVKIMQ-----VQASASSAPNKNQLDGKKKTPSPVKPIPQNTYR 1786
Db 925 RLPPKSDLSLSDSDDDCNLLSQALAGSC-----RPOPSGA-----964
QY 1787 TRVRNADSKNNLNAERVFSDKSKQNKNNKSKDFNDKLPNNEDRVRGSPAFDPSPHY 1846
Db 965 -----STSSSLANASTSTLCRENGQSKQ-----VEHGDK-PN-----996
QY 1847 TPIEGTPCFNRNDSLSLDFDDDDVLSRK-----AELKAKENKESEA-----1992
Db 997 -----YSDDSL-----DDDDDDARSKSLFEQCILSGMHKSNDALESEGEPPGQR 1041
QY 1893 -KVTSHTLTGNO 1904
Db 1042 QEISARDFVSNQ 1054
RESULT 14
ID Q9P119 PRELIMINARY; PRT; 208 AA.
AC Q9P119,
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Adenomatosis polyposis coli tumor suppressor (fragment).
GN APC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435068; PubMed=10982189;
RA Su L.K., Steinbach G., Sawyer J.C., Hindi M., Ward P.A., Lynch P.M.;
RT "Genomic rearrangements of the APC tumor-suppressor gene in familial
RT adenomatous polyposis";
RL Hum. Genet. 106:101-107(2000).
DR EMBL; AF127506; AAF34355.1; -;
DR EMBL; AF127034; AAF34355.1; JOINED.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR00225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 2.
FT NON_TER 1
FT NON_TER 208
SQ SEQUENCE 208 AA; 22726 MW; 51076F257B61C528 CRC64;
Query Match 7.1%; Score 1041; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 5.4e-44;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 471 GLQAIALLQVDCMYGLTNDHYSITLRYAGMALTNLTFGVANKATLCSMKGCMRALV 530
Db 1 GLQAIALLQVDCMYGLTNDHYSITLRYAGMALTNLTFGVANKATLCSMKGCMRALV 60
QY 531 AQLKSESDLOQVIASVLRLNWRADVNSKTLREVGSVKALMECALEVKKESTLKSVL 590
Db 61 AQLKSESDLOQVIASVLRLNWRADVNSKTLREVGSVKALMECALEVKKESTLKSVL 120
QY 591 ALWNLSAHTENKADICAVDGLAFVLGTLTYRSQNTLAIIESGGILRNVSLLIATNE 650
Db 121 ALWNLSAHTENKADICAVDGLAFVLGTLTYRSQNTLAIIESGGILRNVSLLIATNE 180
QY 651 DHRQILRENNCLOTLLQHLKSHSLTIVS 678
Db 181 DHRQILRENNCLOTLLQHLKSHSLTIVS 208
RESULT 15
ID Q72208 PRELIMINARY; PRT; 159 AA.
AC Q72208;

DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Pange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalilus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056268; AAF56268.1; -;
KW Hypothetical protein.
FT NON_TER 159
SQ SEQUENCE 159 AA; 17843 MW; E0C0CC055A22C91B CRC64;
Query Match 5.2%; Score 756; DB 4; Length 159;
Best Local Similarity 98.7%; Pred. No. 4.5e-30;
Matches 153; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAASYDQLLKQVEALKVENSNLKQLEDNSNHLTKLETSNNKEVLKQGSIEDEAM 60
Db 1 MAAASYDQLLKQVEALKVENSNLKQLEDNSNHLTKLETSNNKEVLKQGSIEDEAM 60
QY 61 ASSQIDLLRLKELNLDSSNFPFGVKLRSMKSLRSYSGREGSVSSRGCSVPVMSGFP 120
Db 61 ASSQIDLLRLKELNLDSSNFPFGVKLRSMKSLRSYSGREGSVSSRGCSVPVMSGFP 120
QY 121 RGFVNGSRESTGYLLELEKERSLLLDLDKEEK 155
Db 121 RGFVNGSRESTGYLLELEKERSLLLDLDKEKKK 155
Search completed: August 25, 2004, 17:22:36
Job time : 218.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:01 ; Search time 180.5 Seconds

(without alignments)
4450.319 Million cell updates/sec

Title: US-09-442-489F-2

Perfect score: 14575

Sequence: 1 MAASYPQLKQVZALKWEN.....ESSGTQSPKHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14575	100.0	2843	2 AAR26052	Aar26052 APC gene
2	14575	100.0	2843	2 AAW35392	Aaw35392 Human ade
3	14575	100.0	2843	2 AAW38370	Aaw38370 Human ade
4	14575	100.0	2843	5 ABG90964	Abg90964 Human ade
5	14551	99.8	2860	2 AAR63507	Aar63507 Adenomat
6	14533	99.7	2843	2 AAW76140	Aaw76140 Human APC
7	14533	99.7	2843	2 AAW76144	Aaw76144 Human APC
8	14533	99.7	2843	3 AAB23011	Aab23011 Human APC
9	14533	99.7	2843	5 ABG71105	Abg71105 Human ade
10	14533	99.7	2973	2 AAW76821	Aaw76821 Human APC
11	14533	99.7	2973	4 AAY72782	Aay72782 Transcrip
12	14526	99.7	2973	3 AAY70304	Aay70304 Protein u
13	14521	99.6	2843	2 AAR58634	Aar58634 Adenomat
14	14515.5	99.6	2842	2 AAR63508	Aar63508 Adenomat
15	14515.5	99.6	2842	5 ABG90968	Abg90968 Human APC
16	14515.5	99.6	2843	2 AAW11922	Aaw11922 Adenomat
17	14506	99.5	2843	7 ADE65846	Ade65846 Human ade
18	13966.5	95.8	2742	3 AAB23012	Aab23012 Human APC
19	13168.5	90.3	2842	7 ADS56175	Ads56175 Rat Prote
20	4719	32.4	912	5 ABG71106	Abg71106 Human ade
21	3889	26.7	767	5 ABG71107	Abg71107 Human ade
22	3508.5	24.1	2274	4 AAB50674	Aab50674 Mouse APC
23	3476.5	23.9	902	4 ABG09335	Abg09335 Novel hum
24	3414.5	23.4	2303	6 ABR58648	Abr58648 Human can
25	2171.5	14.9	799	3 AAY92061	Aay92061 Human APC

26	1767.5	12.1	1674	3 AAY92060	Aay92060 Murine AP
27	1763	12.1	2417	4 ABB58126	Abb58126 Drosophil
28	1478	10.1	332	2 AAR88353	Aar88353 Mutant A-
29	1221.5	8.4	1067	4 ABB62156	Abb62156 Drosophil
30	556.5	3.8	2344	4 AAU37120	Aau37120 Staphyloc
31	539	3.7	2768	4 ABB68397	Abb68397 Drosophil
32	526.5	3.6	2586	4 ABB66878	Abb66878 Drosophil
33	524.5	3.6	3257	4 ABB67502	Abb67502 Drosophil
34	518	3.6	2271	6 ABUL6000	Abul6000 Protein e
35	517.5	3.6	2271	6 ABM72734	Abm72734 Staphyloc
36	517.5	3.6	2283	6 ABP56876	Abp56876 Staphyloc
37	516.5	3.5	1186	4 AAB50654	Aab50654 C. elegan
38	514.5	3.5	2261	6 ABJ18914	Abj18914 Pathogen
39	507.5	3.5	178	2 AAW33894	Aaw33894 Flea sal
40	507.5	3.5	178	2 AAW82368	Aaw82368 Flea sal
41	493.5	3.4	2137	5 ABP39618	Abp39618 Staphyloc
42	490	3.4	3111	4 ABB60327	Abb60327 Drosophil
43	481.5	3.3	2951	4 ABB60291	Abb60291 Drosophil
44	474.5	3.3	5533	4 ABB65772	Abb65772 Drosophil
45	474.5	3.3	5560	4 ABB71160	Abb71160 Drosophil

ALIGNMENTS

RESULT 1
AAR26052
ID AAR26052 standard; protein; 2843 AA.
XX AC AAR26052;
XX DT 25-MAR-2003 (revised)
DT 28-JAN-1993 (first entry)
XX XX
DE APC gene product in familial adenomatous polyposis.
XX
XX neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;
XX prognosis; treatment; sporadic colorectal carcinomas; ss.
XX Homo sapiens.
XX WO9213103-A1.
XX PD 06-AUG-1992.
XX PF 16-JAN-1992; 92WO-US000376.
XX PR 16-JAN-1991; 91GB-00000963.
XX PR 08-AUG-1991; 91US-00741940.
XX FA (UYJO) UNIV JOHNS HOPKINS.
XX FA (ICIL) IMPERIAL CHEM IND PLC.
XX FA (UTAH) UNIV UTAH.
XX FA (CANC-) CANCER INST.
XX PI Kinzler KW, Vogelstein B, Anand R, Hedge PJ, Markham AF;
PI Albertsen H, Carlson ML, Groden JL, Joslyn G, Thliveris A, White RL;
PI Nakamura Y;
XX WPI; 1992-284685/34.
XX N-PSDB; AAQ27234.
PT Detection of somatic and germ-line alterations of human APC gene - used
PT to diagnose, treat and study familial adenomatous polyposis and sporadic
PT colorectal cancer.
XX Disclosure; Page 47; 132pp; English.
XX This sequence is encoded by the APC (Adenomatous Polyposis Coli) gene
XX associated with tumorigenesis, found on chromosome 5q. The sequence may
XX be mutated by deletions, insertions, inversions, or point mutations of the
XX gene. The APC gene is expressed in most normal tissues as well suggesting
XX that APC is a tumour suppressor. (Updated on 25-MAR-2003 to correct EN

CC	field.)	(Updated on 25-MAR-2003 to correct PI field.)
XX	Sequence 2843 AA;	
SQ	Query Match	100.0%; Score 14575; DB 2; Length 2843;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 2843; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MAAASYDQLLKQVEALKVENSNTLQFLEEDNSNHLTKLETEASNMKVYLKQLQGSIEDEAM 60
DB	1	MAAASYDQLLKQVEALKVENSNTLQFLEEDNSNHLTKLETEASNMKVYLKQLQGSIEDEAM 60
QY	61	ASSGOIDLLRLKELNLDSSNFPGVKULRSKMSULRSYSGRSVSSRGSCSPVPMGSPFR 120
DB	61	ASSGOIDLLRLKELNLDSSNFPGVKULRSKMSULRSYSGRSVSSRGSCSPVPMGSPFR 120
QY	121	RGFVNGSRESTGYLLEEKERSLILLADLKEEKEKWYIAQLQNLTKRIDSLPTENFSL 180
DB	121	RGFVNGSRESTGYLLEEKERSLILLADLKEEKEKWYIAQLQNLTKRIDSLPTENFSL 180
QY	181	QTLTLRQLRYEARQIRVAMEEQIGTQDMKRAQRRIARIQQIEKDIILRIQLLQSQAT 240
DB	181	QTLTLRQLRYEARQIRVAMEEQIGTQDMKRAQRRIARIQQIEKDIILRIQLLQSQAT 240
QY	241	EAERSSQNKHETGSHDAERQNEGQGVGEINMATSNGQGGSTTRMDHETASVLSSSSTHSA 300
DB	241	EAERSSQNKHETGSHDAERQNEGQGVGEINMATSNGQGGSTTRMDHETASVLSSSSTHSA 300
QY	301	PRRLTSHLGTKVEMYVYLLSMLGTHDKDDMSRTLWASSQDSCISNRQSGCPLLIQLL 360
DB	301	PRRLTSHLGTKVEMYVYLLSMLGTHDKDDMSRTLWASSQDSCISNRQSGCPLLIQLL 360
QY	361	HGNDKDSVLLGNRGSKEARASAAALHNIHSPQDDKRGRRRIIRVLHLEEQIRAYCETC 420
DB	361	HGNDKDSVLLGNRGSKEARASAAALHNIHSPQDDKRGRRRIIRVLHLEEQIRAYCETC 420
QY	421	WEMQEAHEPGWDQKMPMPAPVEHQICPAVCVLMKLSFDEHHRAMNELGGLQAIABELLQ 480
DB	421	WEMQEAHEPGWDQKMPMPAPVEHQICPAVCVLMKLSFDEHHRAMNELGGLQAIABELLQ 480
QY	481	VDCEMTGLTNHYSITILRRYAGALTNLTGADVANKATILCSMGCCRRAVLAQILKSSEDL 540
DB	481	VDCEMTGLTNHYSITILRRYAGALTNLTGADVANKATILCSMGCCRRAVLAQILKSSEDL 540
QY	541	QQVIASVLRNLSWRADVNSKTLREVSGVKALMECALEVKESTLKSVALWNLSAHCT 600
DB	541	QQVIASVLRNLSWRADVNSKTLREVSGVKALMECALEVKESTLKSVALWNLSAHCT 600
QY	601	ENKADICAVDGALAFVGLTITYRSQNTNLAIIESGGGILRNVSLLIATNEDHQIILRENN 660
DB	601	ENKADICAVDGALAFVGLTITYRSQNTNLAIIESGGGILRNVSLLIATNEDHQIILRENN 660
QY	661	CLQTLQLHLKSHSLITIVSNACGTLWNLARNPKDQALWDMGAVSMLKNLIHSHKHWIAM 720
DB	661	CLQTLQLHLKSHSLITIVSNACGTLWNLARNPKDQALWDMGAVSMLKNLIHSHKHWIAM 720
QY	721	GSAALRLNLMANRPAXYKDIANIMSPGSSVPSLHVRKQKALEABDLAQHLSSETFDNIDNLS 780
DB	721	GSAALRLNLMANRPAXYKDIANIMSPGSSVPSLHVRKQKALEABDLAQHLSSETFDNIDNLS 780
QY	781	PKASHRSKQHKOSLXGDEVFDNRHDDNRSNPNFTGNMTVLSPYLTNTVLPSSSSSRGS 840
DB	781	PKASHRSKQHKOSLXGDEVFDNRHDDNRSNPNFTGNMTVLSPYLTNTVLPSSSSSRGS 840
QY	841	LQSSRSKQSLRERERIGLGNTHPATENPGTSSKRGQLQISTTAAQIAKVMEEVSAIHTS 900
DB	841	LQSSRSKQSLRERERIGLGNTHPATENPGTSSKRGQLQISTTAAQIAKVMEEVSAIHTS 900
QY	901	QEDRSSGSTTELHCVTDERNALRPSAAATHSNYNTFTKSENRRTCSPYAKLEYKRS 960
DB	901	QEDRSSGSTTELHCVTDERNALRPSAAATHSNYNTFTKSENRRTCSPYAKLEYKRS 960
QY	961	NDSLNVSNDGVRKQGMKPSTIESYSEDDSFCSGYQVADPLAHKIHSAHMDNDNDE 1020

Db 2041 ECISAMPKKKPKRLKGDNEKHSFPRNMGITLGDLTLDLKDQRPDSEHGLSPDSENF 2100
Qy 2101 WKATQEGANSIVSSILHQAARACLSQAASSDSDLSIKSGISLGSPPHLPDQEEKPFT 2160
Db 2101 WKATQEGANSIVSSILHQAARACLSQAASSDSDLSIKSGISLGSPPHLPDQEEKPFT 2160
Qy 2161 SNKGPRILKPEKSTLETKKIESKSGKIGKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2220
Db 2161 SNKGPRILKPEKSTLETKKIESKSGKIGKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2220
Qy 2221 MPSTSRGTMHIHPCVRNNSSTSPVSKKGPPLKTPAKSKSPSEGOTATTSPRGAKPVKS 2280
Db 2221 MPSTSRGTMHIHPCVRNNSSTSPVSKKGPPLKTPAKSKSPSEGOTATTSPRGAKPVKS 2280
Qy 2281 ELSPVARTQISGSSKAPSGSGRSDSTSPRPAQPLSRPIQSPCRNSISPGRNGISPPN 2340
Db 2281 ELSPVARTQISGSSKAPSGSGRSDSTSPRPAQPLSRPIQSPCRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMQSQNLTKGTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMQSQNLTKGTGLSKNASSIPRSESASKG 2400
Qy 2401 LNMQNGANGKVELSRVSTKSGSGSDRSPVLRQSTFIKEAPSPTRLRKLEESA 2460
Db 2401 LNMQNGANGKVELSRVSTKSGSGSDRSPVLRQSTFIKEAPSPTRLRKLEESA 2460
Qy 2461 SPESLSPSSRPASPTRSOAQTPVLSPSLPDMSLSTHSSVQAGGWRKLPPLSPRIEYNDG 2520
Db 2461 SPESLSPSSRPASPTRSOAQTPVLSPSLPDMSLSTHSSVQAGGWRKLPPLSPRIEYNDG 2520
Qy 2521 RPAKHDIARSHSESPRLPNRSTGTRKSHSSSLPRVSTWERTGSSSSILSASSES 2580
Db 2521 RPAKHDIARSHSESPRLPNRSTGTRKSHSSSLPRVSTWERTGSSSSILSASSES 2580
Qy 2581 SEKAKSEDEKHVNSISGKQKENQVSAKGTWRKIKENEFPTNSTGTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGKQKENQVSAKGTWRKIKENEFPTNSTGTVSSGATNGAES 2640
Qy 2641 KTLIYOMAPAVSKTDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKAMPNIKDSKDN 2700
Db 2641 KTLIYOMAPAVSKTDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKAMPNIKDSKDN 2700
Qy 2701 QAKQNVGNGSVPMRTVGLNRLTISFIQVADPDQKTEIKPGQNNPVPVSETNESPIVERT 2760
Db 2701 QAKQNVGNGSVPMRTVGLNRLTISFIQVADPDQKTEIKPGQNNPVPVSETNESPIVERT 2760
Qy 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPSPRKSSADTSARPSQIPTPVNNTKRDST 2820
Db 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPSPRKSSADTSARPSQIPTPVNNTKRDST 2820
Qy 2821 DSTESSGTQSPKRHSGSYLVTSV 2843
Db 2821 DSTESSGTQSPKRHSGSYLVTSV 2843

RESULT 2

AAW35392

ID AAW35392 standard; protein; 2843 AA.

XX AAW35392;

AC AAW35392;

XX 25-MAR-2003 (revised)

DT 11-MAR-1998 (first entry)

XX Human adenomatous Polyposis coli gene product.

DE Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;

XX Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;

KW neoplastic tissue; tumour tissue; tumour repressor; mutation;

XX sporadic colorectal cancer; detection.

XX Homo sapiens.

OS Homo sapiens.

XX US5648212-A.

PN

XX 15-JUL-1997.
PD
XX 12-AUG-1994; 94US-00289548.
PF
XX 16-JAN-1991; 91GB-00000962.
PR 16-JAN-1991; 91GB-00000963.
PR 16-JAN-1991; 91GB-00000974.
PR 16-JAN-1991; 91GB-00000975.
PR 08-AUG-1991; 91US-00741940.
XX (NICA-) JAPANESE FOUND CANCER RES.
PA (UTAH) UNIV UTAH.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (ZENE) ZENECA LTD.
XX Markham A, Nakamura Y, Groden J, Carlson M, Kinzler K;
PI Albertsen H, Hedge PJ, Vogelstein B, Thliveris A, Anand R, White RL;
PI Joslyn G;
XX WPI; 1997-372053/34.
DR N-PSDB; AAT95538.
XX Cancer diagnosis - by detecting mutation(s) in adenomatous polyposis coli gene.
PT
XX
PS Example 1; Col 33-52; 140pp; English.
XX
CC The present sequence is the human adenomatous Polyposis coli (APC) gene product, which was used in the development of a novel method of diagnosing or prognosing an APC gene associated neoplastic tissue. The method comprises comparing an APC gene coding sequences or mRNA in a tumour tissue, to APC gene coding sequences or mRNA in a non-neoplastic tissue, where a difference indicates an APC gene associated neoplasia of the tumour tissue. APC is a tumour repressor expressed in most normal tissues. APC mutations are found in familial adenomatous polyposis and sporadic colorectal cancer patients. The method enables mutations to be detected to provide an indication of predisposition to cancer. (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2843 AA;

Query Match 100.0%; Score 14575; DB 2; Length 2843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAAASYDOLLKQVEALKVENSRLRQLEDNSNHLTKLETEASNKVEYLKOLQSGIDEAM 60
Db 1 MAAASYDOLLKQVEALKVENSRLRQLEDNSNHLTKLETEASNKVEYLKOLQSGIDEAM 60
Qy 61 ASSQIDLLERLKLNDSSNFFGVKLRKMSLRYSYSGREGSVSSRSGECSPPVPMGSFPR 120
Db 61 ASSQIDLLERLKLNDSSNFFGVKLRKMSLRYSYSGREGSVSSRSGECSPPVPMGSFPR 120
Qy 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKDYAOLNLTGRIDSLPTENFSL 180
Db 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKDYAOLNLTGRIDSLPTENFSL 180
Qy 181 QTDLTRQLEYEARQIRVAMEEQIGTQDMKEKAQRRIARIQIQIEKDILRIQLQSQAT 240
Db 181 QTDLTRQLEYEARQIRVAMEEQIGTQDMKEKAQRRIARIQIQIEKDILRIQLQSQAT 240
Qy 241 EAERSSQNKHETGSHDAERONEGQGVGEINATSGNCGQSTTRMDHETASVSSSTHSA 300
Db 241 EAERSSQNKHETGSHDAERONEGQGVGEINATSGNCGQSTTRMDHETASVSSSTHSA 300
Qy 301 PRLTSHLGTKEVMYVLSLLMLGTHDKDDMSRTLLAMSSQDSCISMROSGCPLLIQLL 360
Db 301 PRLTSHLGTKEVMYVLSLLMLGTHDKDDMSRTLLAMSSQDSCISMROSGCPLLIQLL 360
Qy 361 HGNDKDSVLLGNSRSGSEARASAAALHNIHSGOPDKRGRREIRVHLLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRSGSEARASAAALHNIHSGOPDKRGRREIRVHLLEQIRAYCETC 420

QY 421 WENQEAHEPGMDQDNKPNPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGGLQAIABELLQ 480
 Db 421 WENQEAHEPGMDQDNKPNPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGGLQAIABELLQ 480
 QY 481 VDCMYGLTNDHYITILRRVAGMALTNLTGFDVANKATLCSMGKCMALVAQLKSESED 540
 Db 481 VDCMYGLTNDHYITILRRVAGMALTNLTGFDVANKATLCSMGKCMALVAQLKSESED 540
 QY 541 QOVTASVLRNLWSRADVNSKKTLEVGSGVKALMECALVEKSTLSVLSALWNLSAHCT 600
 Db 541 QOVTASVLRNLWSRADVNSKKTLEVGSGVKALMECALVEKSTLSVLSALWNLSAHCT 600
 QY 601 ENKADIICAVDICALAFVGTLYTYSQTNWTLAIIESGGGILNVSSLLIATNEDHRCIIRENN 660
 Db 601 ENKADIICAVDICALAFVGTLYTYSQTNWTLAIIESGGGILNVSSLLIATNEDHRCIIRENN 660
 QY 661 CLOTLLOHLKSHSLTIVSNACGTLWNLARNPKDQOEALWDMGAVSMLKNLIHSHKHVIAM 720
 Db 661 CLOTLLOHLKSHSLTIVSNACGTLWNLARNPKDQOEALWDMGAVSMLKNLIHSHKHVIAM 720
 QY 721 GSAALRNLANRPAKYKDANINSPGSSLSPLHVRKQKALEAEIDACHILSETFDNDNLS 780
 Db 721 GSAALRNLANRPAKYKDANINSPGSSLSPLHVRKQKALEAEIDACHILSETFDNDNLS 780
 QY 781 PKASHRSKQRHKQSLYGDYVFTNRRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSSRGS 840
 Db 781 PKASHRSKQRHKQSLYGDYVFTNRRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSSRGS 840
 QY 841 LDSRSSEKDSRLERERGIGLGNHPATENPCTSKRGLOISTTAAQIAKWMESVSAIHVS 900
 Db 841 LDSRSSEKDSRLERERGIGLGNHPATENPCTSKRGLOISTTAAQIAKWMESVSAIHVS 900
 QY 901 QEDSSSGSTTELHCVTDERNALRRSSAAHSTNFTKSENTRCSPYAKLEYKRGS 960
 Db 901 QEDSSSGSTTELHCVTDERNALRRSSAAHSTNFTKSENTRCSPYAKLEYKRGS 960
 QY 961 NDSLNSVSSNDGYKQOMKPSIESYSEDESXFCSGYQYPAOLAUKIHSANHMDNDGE 1020
 Db 961 NDSLNSVSSNDGYKQOMKPSIESYSEDESXFCSGYQYPAOLAUKIHSANHMDNDGE 1020
 QY 1021 LDTPIYSLKYDEQLNSGRQSPONERWARPKHIIIDEIKQSEQRQSNQSTTYPVYTE 1080
 Db 1021 LDTPIYSLKYDEQLNSGRQSPONERWARPKHIIIDEIKQSEQRQSNQSTTYPVYTE 1080
 QY 1081 STDDKHLKFPQHFQOQECVSPYRSRGANGSETNVRVGSNHGINQVNSQSLCQEDDYEDDKP 1140
 Db 1081 STDDKHLKFPQHFQOQECVSPYRSRGANGSETNVRVGSNHGINQVNSQSLCQEDDYEDDKP 1140
 QY 1141 TNSYERYSEOEHEEREPNYSIKYNEEKRHVDQPIDYSLKYATDIBSOKQSFSSKS 1200
 Db 1141 TNSYERYSEOEHEEREPNYSIKYNEEKRHVDQPIDYSLKYATDIBSOKQSFSSKS 1200
 QY 1201 SSGQSSKTEHWSSESSENTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
 Db 1201 SSGQSSKTEHWSSESSENTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
 QY 1261 TYCVEDTPICFRSCSSLSLSAEDTGCNQTQEADSNANTLOIAIKIGITRSNEDPV 1320
 Db 1261 TYCVEDTPICFRSCSSLSLSAEDTGCNQTQEADSNANTLOIAIKIGITRSNEDPV 1320
 QY 1321 SEVPAVSQHPRTKSSRLQSSLSSESARHKAVEFPSCAKSPSKSGAQTPKSPPEHYVQET 1380
 Db 1321 SEVPAVSQHPRTKSSRLQSSLSSESARHKAVEFPSCAKSPSKSGAQTPKSPPEHYVQET 1380
 QY 1381 PLMFSTRCTSVSSLDSFESRSIASVQSEPCSGMVSGIISPSDLDPSPGOTMPPPSRSKTPP 1440
 Db 1381 PLMFSTRCTSVSSLDSFESRSIASVQSEPCSGMVSGIISPSDLDPSPGOTMPPPSRSKTPP 1440
 QY 1441 PPTQATQKEVPKNKAPTAKESGPKQAAVNAVQVLPDADTLHLPATSTPDGF 1500
 Db 1441 PPTQATQKEVPKNKAPTAKESGPKQAAVNAVQVLPDADTLHLPATSTPDGF 1500

QY 1501 SCSSLSALSLEDPPIQKDVLRIMPVQENDNGNETESEBQKESNENQEKAEKTTDSE 1560
 Db 1501 SCSSLSALSLEDPPIQKDVLRIMPVQENDNGNETESEBQKESNENQEKAEKTTDSE 1560
 QY 1561 KOLLDDSDDDDIIEILEECIIISAMPTKSRKKGKPAQATASKLPPPVARKPSQLPYVKLLPS 1620
 Db 1561 KOLLDDSDDDDIIEILEECIIISAMPTKSRKKGKPAQATASKLPPPVARKPSQLPYVKLLPS 1620
 QY 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIPNFSTATSLDITIESPPNELAAGEGVGGAQ 1680
 Db 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIPNFSTATSLDITIESPPNELAAGEGVGGAQ 1680
 QY 1681 SGEPEKRTIITEGRSTDEAOGKTSVVTIPELDNKAEEGDILAEICINSAMPKKGSHKP 1740
 Db 1681 SGEPEKRTIITEGRSTDEAOGKTSVVTIPELDNKAEEGDILAEICINSAMPKKGSHKP 1740
 QY 1741 FRVKKIMDQVOQAASASSAPNKNQLDGKKKXPTSPVKPIPONTYRTRVRKNADSKNNLN 1800
 Db 1741 FRVKKIMDQVOQAASASSAPNKNQLDGKKKXPTSPVKPIPONTYRTRVRKNADSKNNLN 1800
 QY 1801 AERVFSDNKKSKQNLKNNKDFNDKLPNNEDRVGSGFADSPHHYTPIEGTYPYCFPSRND 1860
 Db 1801 AERVFSDNKKSKQNLKNNKDFNDKLPNNEDRVGSGFADSPHHYTPIEGTYPYCFPSRND 1860
 QY 1861 SLSLDFDDDDVDLSREKAELEKAKENKESAKVTSHTELTSNQOSANKTQAIKQPINR 1920
 Db 1861 SLSLDFDDDDVDLSREKAELEKAKENKESAKVTSHTELTSNQOSANKTQAIKQPINR 1920
 QY 1921 COPRPILQKQSTFPQSSKDIIDRGAANTDEKLNPAIENTPVCFSHNSLSLSLSDIDENN 1980
 Db 1921 COPRPILQKQSTFPQSSKDIIDRGAANTDEKLNPAIENTPVCFSHNSLSLSLSDIDENN 1980
 QY 1981 NKNEPIKETEPDPSQGEPPSPQASGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
 Db 1981 NKNEPIKETEPDPSQGEPPSPQASGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
 QY 2041 ECISAMPKKKKPRLKGDNEKHSFRNMGGLGLEDLTLDKDIQRPDSEHGLSPDSNFD 2100
 Db 2041 ECISAMPKKKKPRLKGDNEKHSFRNMGGLGLEDLTLDKDIQRPDSEHGLSPDSNFD 2100
 QY 2101 WKATOEGANSIVSSLHQAAAACLSROASDSDBSILSKGISLGSPFHLTPDQEKPF 2160
 Db 2101 WKATOEGANSIVSSLHQAAAACLSROASDSDBSILSKGISLGSPFHLTPDQEKPF 2160
 QY 2161 SNKGPRILKPEKSTLETTKITESKGIKGGKVKYKSLITGKVRNSSEISQWKKQPIQAN 2220
 Db 2161 SNKGPRILKPEKSTLETTKITESKGIKGGKVKYKSLITGKVRNSSEISQWKKQPIQAN 2220
 QY 2221 MPSISRGRTMHIIPGVNNSSTSPVSKGPPPLKTPASKSPSEGTATTPRGAKPVS 2280
 Db 2221 MPSISRGRTMHIIPGVNNSSTSPVSKGPPPLKTPASKSPSEGTATTPRGAKPVS 2280
 QY 2281 ELSVPAQTQOIGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGISPPN 2340
 Db 2281 ELSVPAQTQOIGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPGRNSISPRNGISPPN 2340
 QY 2341 KLSOLPRTSPSTASTKSSGSGKMSYTPGRQWQOONLTQOTGLSKNASSIPRSESASK 2400
 Db 2341 KLSOLPRTSPSTASTKSSGSGKMSYTPGRQWQOONLTQOTGLSKNASSIPRSESASK 2400
 QY 2401 LNQMNGANGAKKVELSRMSSTKSSGSESDRSPVLRQSTTIKEAPSTLRRKLEESA 2460
 Db 2401 LNQMNGANGAKKVELSRMSSTKSSGSESDRSPVLRQSTTIKEAPSTLRRKLEESA 2460
 QY 2461 SFESLSRSPRSPATRSQAOTPVLSPLDMSLTHSSVQAGWRKLPNLSLTIEYNDG 2520
 Db 2461 SFESLSRSPRSPATRSQAOTPVLSPLDMSLTHSSVQAGWRKLPNLSLTIEYNDG 2520
 QY 2521 RPAKRHDIAKSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTTGSSSSILSASSES 2580
 Db 2521 RPAKRHDIAKSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTTGSSSSILSASSES 2580
 QY 2581 SEKAKSEDEKRVNSISGTQKOSKNQVSAKTWTKIKENEFSPNITSQTIVSSGATNGAES 2640

Db 2581 SEKAKSEDEKRVNSISGTSKQENQVSAKGTWRKIKENEFPSTNSTQTVSSGATNGAES 2640
Qy 2641 KTLIIQWAPAVSKTDEWVRIEDCPINNPRSGRSPGTNPVVIDSVSEKAPNPKDSKDN 2700
Db 2641 KTLIIQWAPAVSKTDEWVRIEDCPINNPRSGRSPGTNPVVIDSVSEKAPNPKDSKDN 2700
Qy 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVDAPDQKGTBIKPGQNNPVPVSTNESPVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLTSFIQVDAPDQKGTBIKPGQNNPVPVSTNESPVERT 2760
Qy 2761 PFSSSSSKHSSPSGTVAAARVTPNNYNPSPRKSADSTARSPOIPTPVNNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAAARVTPNNYNPSPRKSADSTARSPOIPTPVNNNTKKRDSKT 2820
Qy 2821 DSTESSGTQSPKXHSGLYVTSV 2843
Db 2821 DSTESSGTQSPKXHSGLYVTSV 2843
RESULT 3
AAW38370
ID AAW38370 standard; protein; 2843 AA.
XX
AC AAW38370;
XX
DT 25-MAR-2003 (revised)
DT 08-APR-1998 (first entry)
XX
DE Human adenomatous Polyposis coli gene product.
XX
KW Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;
KW neoplastic tissue; tumour tissue; tumour repressor; mutation;
KW sporadic colorectal cancer; detection.
XX
OS Homo sapiens.
XX
PN US5691454-A.
XX
PD 25-NOV-1997.
XX
PF 25-MAY-1995; 95US-00452854.
XX
PR 16-JAN-1991; 91GB-00000962.
PR 16-JAN-1991; 91GB-00000963.
PR 16-JAN-1991; 91GB-00000974.
PR 16-JAN-1991; 91GB-00000975.
PR 08-AUG-1991; 91US-00741340.
PR 12-AUG-1994; 94US-00289548.
XX
PA (CANC-) CANCER INST.
PA (UJJO) UNIV JOHNS HOPKINS.
PA (ICIL) IMPERIAL CHEM IND PLC.
PA (UTAH) UNIV UTAH.
XX
Pi Nakamura Y, Markham AP, Groden J, Vogelstein B, Kinzler K;
Pi Hedge PJ, Carlson M, Anand R, Thliveris A, Albertsen H, White RL;
Pi Joslyn G;
XX
DR WPI: 1998-017712/02.
DR N-P8DB; AAT96153.
XX
PT Antibodies to normal and mutant adenomatous polyposis coli proteins -
PT useful for detecting genetic predisposition to cancer.
XX
PS Example 1; Col 33-52; 107pp; English.
XX
CC The present sequence is the human adenomatous Polyposis coli (APC) gene
CC product, which was used in the development of a novel method of
CC diagnosing or prognosing an APC gene associated neoplastic tissue. The
CC method comprises comparing APC gene coding sequences or mRNA in a tumour
CC tissue, to APC gene coding sequences or mRNA in a non-neoplastic tissue,
CC where a difference indicates an APC gene associated neoplasia of the

CC tumour tissue. APC is a tumour repressor expressed in most normal
CC tissues. APC mutations are found in familial adenomatous polyposis and
CC sporadic colorectal cancer patients. The method enables mutations to be
CC detected to provide an indication of predisposition to cancer. (Updated
CC on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2843 AA;
Query Match 100.0%; Score 14575; DB 2; Length 2843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAASVDOLLKQVEALKWENSLROLEDNSNHLTKLEASNMKEVLKQLOGSTEDAM 60
Db 1 MAASVDOLLKQVEALKWENSLROLEDNSNHLTKLEASNMKEVLKQLOGSTEDAM 60
Qy 61 ASSGQIDLLERLKEINLDSSNPPGPKLRSKWSLRSYGSRGSSVSSRSGECSPVPMGSPFR 120
Db 61 ASSGQIDLLERLKEINLDSSNPPGPKLRSKWSLRSYGSRGSSVSSRSGECSPVPMGSPFR 120
Qy 121 RGFVNGSRSTCYLEBELEKERSLLADLDKEEKDWTYYAQLONLTKRIDSPLTENFSL 180
Db 121 RGFVNGSRSTCYLEBELEKERSLLADLDKEEKDWTYYAQLONLTKRIDSPLTENFSL 180
Qy 181 QTDLTRQLEYEARIQIRVAMEBOLGTCDMEKRAQRRIARIQOIEKIDILRIQLQSQAT 240
Db 181 QTDLTRQLEYEARIQIRVAMEBOLGTCDMEKRAQRRIARIQOIEKIDILRIQLQSQAT 240
Qy 241 EAERSQNKHETGSHDAERQNEGGVGEINMATSGNGGSTRMDHETASVLSSTSTSA 300
Db 241 EAERSQNKHETGSHDAERQNEGGVGEINMATSGNGGSTRMDHETASVLSSTSTSA 300
Qy 301 PRLTSHLGTKEVMYVYLLSLMGLTHDKDMSRTLLAMSSQSCISMRQSGCLPLLIQLL 360
Db 301 PRLTSHLGTKEVMYVYLLSLMGLTHDKDMSRTLLAMSSQSCISMRQSGCLPLLIQLL 360
Qy 361 HGNKDSVLLGNSRGSKEARAPASAAALHNI IHSQDDKRGREIRVLLHLEQIRAYCETC 420
Db 361 HGNKDSVLLGNSRGSKEARAPASAAALHNI IHSQDDKRGREIRVLLHLEQIRAYCETC 420
Qy 421 WEQBAHEPGMDQDNPMAPVPEHQICPAVCVLMKLSFDEEHRHANNELGGQATLQLL 480
Db 421 WEQBAHEPGMDQDNPMAPVPEHQICPAVCVLMKLSFDEEHRHANNELGGQATLQLL 480
Qy 481 VDCEMYGLTNDHYSITLRRYAGVALTNTLTFDGVANKATLCSMKGMWALVAQLKSEDEL 540
Db 481 VDCEMYGLTNDHYSITLRRYAGVALTNTLTFDGVANKATLCSMKGMWALVAQLKSEDEL 540
Qy 541 QQVIASVLENLNLSWRADVNSKTLREVGSVKALMECALEVKKESTILKSVLSALWNLNLSAHT 600
Db 541 QQVIASVLENLNLSWRADVNSKTLREVGSVKALMECALEVKKESTILKSVLSALWNLNLSAHT 600
Qy 601 ENKADICAVDGLAFLVGLTYRSOTNTLAIIESGGILRNVSSTIATNEDHRQLRENN 660
Db 601 ENKADICAVDGLAFLVGLTYRSOTNTLAIIESGGILRNVSSTIATNEDHRQLRENN 660
Qy 661 CLQTLQHLKSHSLTIVSNACGTLWNLNLSARNPKDQEALWMDGAVSNLKNL IHSKHMIAM 720
Db 661 CLQTLQHLKSHSLTIVSNACGTLWNLNLSARNPKDQEALWMDGAVSNLKNL IHSKHMIAM 720
Qy 721 GSAALRNLMANRPAYKIDANIMSPGSSILPSLHVKKOKALBAELDAQHLSETFDNIDLS 780
Db 721 GSAALRNLMANRPAYKIDANIMSPGSSILPSLHVKKOKALBAELDAQHLSETFDNIDLS 780
Qy 781 PKASHRSKORHKQSLYGDYVFTDNRHDDNRSDFNTGNMTVLSPLYNTTLPSSSSSRGS 840
Db 781 PKASHRSKORHKQSLYGDYVFTDNRHDDNRSDFNTGNMTVLSPLYNTTLPSSSSSRGS 840
Qy 841 LPSRSRSEKDRSLEREREGIGLGNHYHATENPGTSSKRGLOISTAAQIAKMWEEVAIHTS 900
Db 841 LPSRSRSEKDRSLEREREGIGLGNHYHATENPGTSSKRGLOISTAAQIAKMWEEVAIHTS 900
Qy 901 QEDRSSGSGTTELHCVTDERNALRRSSAAHTSNTYNTFKSENSNRTCSMPYAKLEYKRSS 960

Db 901 QEDRSSGSGTTELHCVTDERNALRRSAAHTNTYNTFTKSENSNRTCSMPYAKLYKRS 960
Qy 961 NDSLNSVSSNDGCGKQGMKPSIESYSEDESKFCSYGOYPADLAHKIHSANMMDNDGE 1020
Db 961 NDSLNSVSSNDGCGKQGMKPSIESYSEDESKFCSYGOYPADLAHKIHSANMMDNDGE 1020
Qy 1021 LDTPIINYSIKYDEQLNSGRQSPSONERWARPKHIIIEDEIKOSEQRQSNQSTTVPVYTE 1080
Db 1021 LDTPIINYSIKYDEQLNSGRQSPSONERWARPKHIIIEDEIKOSEQRQSNQSTTVPVYTE 1080
Qy 1081 STDDXHLKQPHFGQCECVSPYRSRGANGSETRVGSNGINONVSQSLCQEDDYEEDKP 1140
Db 1081 STDDXHLKQPHFGQCECVSPYRSRGANGSETRVGSNGINONVSQSLCQEDDYEEDKP 1140
Qy 1141 TNYSEYSEEEQHEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSSKS 1200
Db 1141 TNYSEYSEEEQHEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSSKS 1200
Qy 1201 SSGQSKTHMSSSSSENTSTPSSNAKRONOLHPSSAQSRSGOPKAACTKUSSINOETIQ 1260
Db 1201 SSGQSKTHMSSSSSENTSTPSSNAKRONOLHPSSAQSRSGOPKAACTKUSSINOETIQ 1260
Qy 1261 TYCVEDTPICFGRCSLSLSAEDIEGNCQTTQEDADSANTLIQIAEIKIGKIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFGRCSLSLSAEDIEGNCQTTQEDADSANTLIQIAEIKIGKIGTRSAEDPV 1320
Qy 1321 SEVPAVSQHPRTKSSRLQSSLSSESARHKAVEFFSGAKSPKSGAOTPKSPPEHYVQET 1380
Db 1321 SEVPAVSQHPRTKSSRLQSSLSSESARHKAVEFFSGAKSPKSGAOTPKSPPEHYVQET 1380
Qy 1381 PLMFSSCTSVSSLDSPESASIASVQSPGCSGVSGIISPSDLPSPGQTMPPSRSKTTP 1440
Db 1381 PLMFSSCTSVSSLDSPESASIASVQSPGCSGVSGIISPSDLPSPGQTMPPSRSKTTP 1440
Qy 1441 PPQTAQTRKVPKNAPTAEKRESGPKQAANAAVQVRQVLPDADTLHLFATESTPDGF 1500
Db 1441 PPQTAQTRKVPKNAPTAEKRESGPKQAANAAVQVRQVLPDADTLHLFATESTPDGF 1500
Qy 1501 SCSSLSALSLEDPFQKQVELRIMPVQENDNGNETSEOPKESENEQKEAETIDSE 1560
Db 1501 SCSSLSALSLEDPFQKQVELRIMPVQENDNGNETSEOPKESENEQKEAETIDSE 1560
Qy 1561 KOLLDSDDDDIEILBECIISAMPTKSRGKPKAQTASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDSDDDDIEILBECIISAMPTKSRGKPKAQTASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNPLQPKHVSPTPGDMPRVYCVGTPINPSTATSLDLTIESPPNELAAGEGVGGQA 1680
Db 1621 QNPLQPKHVSPTPGDMPRVYCVGTPINPSTATSLDLTIESPPNELAAGEGVGGQA 1680
Qy 1681 SGFEFKRDTIPTEGRSTDEAQGKTSSTVITPELDDNKABEGDILAEICINSAMPKGSHP 1740
Db 1681 SGFEFKRDTIPTEGRSTDEAQGKTSSTVITPELDDNKABEGDILAEICINSAMPKGSHP 1740
Qy 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPQNTYRTRVRKNADSNNLN 1800
Db 1741 FRVKKIMDQVQASASSAPNKNQLDGKKKPTSPVKPIPQNTYRTRVRKNADSNNLN 1800
Qy 1801 AERVFSNDKSKQNLKNSKDFNDKLPNNEDRVGSAFDPSPHYTPIEGTPYCFSRND 1860
Db 1801 AERVFSNDKSKQNLKNSKDFNDKLPNNEDRVGSAFDPSPHYTPIEGTPYCFSRND 1860
Qy 1861 SLSLDFDDDDVDLSREKAELEKAKENKSEAKVTSHTELTSNQOSANKTOIAKQPINR 1920
Db 1861 SLSLDFDDDDVDLSREKAELEKAKENKSEAKVTSHTELTSNQOSANKTOIAKQPINR 1920
Qy 1921 GQPKPILOKQSTFPQSSKDIIPDRGAATDEKLQNFALIENTPVCFSHNSLSLSDIDQENN 1980
Db 1921 GQPKPILOKQSTFPQSSKDIIPDRGAATDEKLQNFALIENTPVCFSHNSLSLSDIDQENN 1980
Qy 1981 NKNEPIKETEPPDQSGEPPSKQASGYAPKSFHVEDTPVCFSRNSLSLSIDSEDDLLQ 2040

Db 1981 NKNEPIKETEPPDQSGEPPSKQASGYAPKSFHVEDTPVCFSRNSLSLSIDSEDDLLQ 2040
Qy 2041 ECISSAMPKKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDQIRPDSEHGLSPDSENF 2100
Db 2041 ECISSAMPKKKPSRLKGDNEKHSPRNMGILGEDLTLDLKDQIRPDSEHGLSPDSENF 2100
Qy 2101 WKAIQEGANSIVSLHQAAAAACLSRQASDSDSILSLKSGISLSPFHLTPDQEKXPT 2160
Db 2101 WKAIQEGANSIVSLHQAAAAACLSRQASDSDSILSLKSGISLSPFHLTPDQEKXPT 2160
Qy 2161 SNKGPRLLKPEGEKSTLETKKIESEKIGKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
Db 2161 SNKGPRLLKPEGEKSTLETKKIESEKIGKGGKVKYKSLITGKVRNSEISGQMKQPLQAN 2220
Qy 2221 MPISIRGRTMIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGGTATTSPRGAKEPVKS 2280
Db 2221 MPISIRGRTMIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGGTATTSPRGAKEPVKS 2280
Qy 2281 ELSFVARQTSQIGSSKAPSRSGSRDSTPSRPAQOPLSRPIOSPGRNSISPGRNGISPPN 2340
Db 2281 ELSFVARQTSQIGSSKAPSRSGSRDSTPSRPAQOPLSRPIOSPGRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSPSTASTKSGSGKVSYSYTSRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSPSTASTKSGSGKVSYSYTSRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
Qy 2401 LNQMNGNGANKKVELSRMSSTKSGSGESDRSERPVLVRQSTFIKEAPSTLRRKLEESA 2460
Db 2401 LNQMNGNGANKKVELSRMSSTKSGSGESDRSERPVLVRQSTFIKEAPSTLRRKLEESA 2460
Qy 2461 SFESLSPSSPASPTRSAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
Db 2461 SFESLSPSSPASPTRSAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
Qy 2521 RPAKHDIASHSESPSRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
Db 2521 RPAKHDIASHSESPSRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSILSASSES 2580
Qy 2581 SEKASEDEKHVANSISGTTQSKENQVSAGTKWIKIKENEFSPNTNSTQTVSSGATNGAES 2640
Db 2581 SEKASEDEKHVANSISGTTQSKENQVSAGTKWIKIKENEFSPNTNSTQTVSSGATNGAES 2640
Qy 2641 KTLIIYQMAPAVSKTEDVWVRIEDCPINNPRSGSPGTNTPPVDSVSEKANPNIKDSKN 2700
Db 2641 KTLIIYQMAPAVSKTEDVWVRIEDCPINNPRSGSPGTNTPPVDSVSEKANPNIKDSKN 2700
Qy 2701 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTETKPGQNNPVVSETNESPIVERT 2760
Db 2701 QAKQNVGNGVPMRTVGLNRLTSFIQVDAPDQKGTETKPGQNNPVVSETNESPIVERT 2760
Qy 2761 PFSSSSSKHSSPSGTVAARVTFNNPNPRSKSADSTARSQIPTPVNNNTKGRDST 2820
Db 2761 PFSSSSSKHSSPSGTVAARVTFNNPNPRSKSADSTARSQIPTPVNNNTKGRDST 2820
Qy 2821 DSTEESGTSQSPKSRHSGSYLVTSV 2843
Db 2821 DSTEESGTSQSPKSRHSGSYLVTSV 2843
RESULT 4
ABG90964
ID ABG90964 standard; protein; 2843 AA.
XX ABG90964;
XX AC ABG90964;
XX XX
DT 29-NOV-2002 (first entry)
XX Human adenomatous polyposis coli (APC) protein #1.
XX Adenomatous polyposis coli; APC; human; neoplastic tissue;
XX mutation detection; tumour; cancer.
OS Homo sapiens.

Db 1441 PPQTAQTRVPPKNAKTAETKRESGPKQAAVNAVQVQLPDAADTLILHPATESTPDGF 1500
 QY 1501 SCSSLSALSLDEPFTQKDVLRIMPPVOENDNGNETSEQPKESNENQOEKAEKTIIDSE 1560
 Db 1501 SCSSLSALSLDEPFTQKDVLRIMPPVOENDNGNETSEQPKESNENQOEKAEKTIIDSE 1560
 QY 1561 KOLLDDSDDDDEILIEECIIISAMPTKSRKGGKPAQTAKLPPVAPKPSQLPVYKLLPS 1620
 Db 1561 KOLLDDSDDDDEILIEECIIISAMPTKSRKGGKPAQTAKLPPVAPKPSQLPVYKLLPS 1620
 QY 1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGEGVGGGAQ 1680
 Db 1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPPNELAAGEGVGGGAQ 1680
 QY 1681 SGFEKRDITPEGRSTDEAQQGKTSSVTIPBLDNKABEGDILAECTINSAMPKGSHPK 1740
 Db 1681 SGFEKRDITPEGRSTDEAQQGKTSSVTIPBLDNKABEGDILAECTINSAMPKGSHPK 1740
 QY 1741 FRVKITMDVQQAASASSAPNKNQDGGKKTSPVKPIQNTYRTRVRKNVADSKNNLN 1800
 Db 1741 FRVKITMDVQQAASASSAPNKNQDGGKKTSPVKPIQNTYRTRVRKNVADSKNNLN 1800
 QY 1801 AERVFSDNDSKQNLKNNKDFNDKLPNNEDRVGSGFAPDSFPHYTYPIEGTPYCFSRND 1860
 Db 1801 AERVFSDNDSKQNLKNNKDFNDKLPNNEDRVGSGFAPDSFPHYTYPIEGTPYCFSRND 1860
 QY 1861 SLSSLDFFDDDDVLSREKAEELKAKENKESAKVTSHTELTSNQQSANKTOAIKAPINR 1920
 Db 1861 SLSSLDFFDDDDVLSREKAEELKAKENKESAKVTSHTELTSNQQSANKTOAIKAPINR 1920
 QY 1921 GQPKPILOKQSTPQSSKDIPIRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 Db 1921 GQPKPILOKQSTPQSSKDIPIRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
 QY 1981 KNEPEIKETEPPDSOGESKQAGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENLLQ 2040
 Db 1981 KNEPEIKETEPPDSOGESKQAGYAPKSFHVEDTPVCFSRNSLSLSLSDIDQENLLQ 2040
 QY 2041 ECITSSAMPKKKPSRLKGNKEHSPRNMGGIIGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 Db 2041 ECITSSAMPKKKPSRLKGNKEHSPRNMGGIIGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 QY 2101 WKAIQEGANSIVSSLHQAACLSQASDSDSLKSLGSLGSPHLPDQEKSPFT 2160
 Db 2101 WKAIQEGANSIVSSLHQAACLSQASDSDSLKSLGSLGSPHLPDQEKSPFT 2160
 QY 2161 SNKGPIILPGEKSTLTETKIESESGIKGGKVKSLITGVRNSLSLSLSDIDQENLLQ 2220
 Db 2161 SNKGPIILPGEKSTLTETKIESESGIKGGKVKSLITGVRNSLSLSLSDIDQENLLQ 2220
 QY 2221 MPSISGRMTIHIPVGRNSSTSPVSKGPKLTPASKSPSEGOTATTPRGAKPVSYS 2280
 Db 2221 MPSISGRMTIHIPVGRNSSTSPVSKGPKLTPASKSPSEGOTATTPRGAKPVSYS 2280
 QY 2281 ELSPVARQTSOIGGSKAPSRGSRDSTPSRAQOPLSRPIQSPGRNSISPGNGISPPN 2340
 Db 2281 ELSPVARQTSOIGGSKAPSRGSRDSTPSRAQOPLSRPIQSPGRNSISPGNGISPPN 2340
 QY 2341 KLSQLPRTSPSTASTKSSGSGKMSYTPSGRQVMSQNLTKOTGLSKNASSIPRSESASKG 2400
 Db 2341 KLSQLPRTSPSTASTKSSGSGKMSYTPSGRQVMSQNLTKOTGLSKNASSIPRSESASKG 2400
 QY 2401 LNMNNGANGANKVELSRMSSTKSSGSDRERPVLRQSTFIKEAPSTLRKLKLESA 2460
 Db 2401 LNMNNGANGANKVELSRMSSTKSSGSDRERPVLRQSTFIKEAPSTLRKLKLESA 2460
 QY 2461 SFESLSPSRPAPTRSOAQTPVLSPLDMSLSTHSSVQAGWWRKLPNLSPTIEYNDG 2520
 Db 2461 SFESLSPSRPAPTRSOAQTPVLSPLDMSLSTHSSVQAGWWRKLPNLSPTIEYNDG 2520
 QY 2521 RPAKRHDIAHSHSPRLPINSPTGWTKEHSGHSSLPVSTWRTGSSSSILSASSES 2580
 Db 2521 RPAKRHDIAHSHSPRLPINSPTGWTKEHSGHSSLPVSTWRTGSSSSILSASSES 2580

QY 2581 SEKAKSEDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSPNTSOTVSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSPNTSOTVSSGATNGAES 2640
 QY 2641 KTLIIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
 QY 2701 QAKQNVGNGVPMRTVGLNRLTSFTQVDAPDQKGTETKPGQNNPVVSETNESPIVET 2760
 Db 2701 QAKQNVGNGVPMRTVGLNRLTSFTQVDAPDQKGTETKPGQNNPVVSETNESPIVET 2760
 QY 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPRKSSADSTGARPSTQPTPVNNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPRKSSADSTGARPSTQPTPVNNNTKKRDSKT 2820
 QY 2821 DSTESSGTSQPKHSGSYLVTSV 2843
 Db 2821 DSTESSGTSQPKHSGSYLVTSV 2843
 RESULT 5
 AAR63507
 ID AAR63507 standard; protein; 2860 AA.
 XX AAR63507;
 AC AC
 XX 25-MAR-2003 (revised)
 DT 30-MAY-1995 (first entry)
 XX
 DE Adenomatous polyposis coli tumour repressor.
 XX
 KW Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;
 KW familial adenomatous polyposis; cancer diagnosis and prognosis;
 KW tumorigenesis suppression.
 XX
 OS Homo sapiens.
 XX
 FN US5352775-A.
 XX
 PD 04-OCT-1994.
 XX
 PF 08-AUG-1991; 91US-00741940.
 XX
 PR 16-JAN-1991; 91GB-00000962.
 PR 16-JAN-1991; 91GB-00000963.
 PR 16-JAN-1991; 91GB-00000974.
 PR 16-JAN-1991; 91GB-00000975.
 XX
 PA (ICIL) IMPERIAL CHEM IND.
 PA (CANC-) CANCER INST.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 XX
 PI Markham AP, Hedge PJ, Anand R, Nakamura Y, Groden J, Kinzler K;
 PI Thliveris A, Carlson M, Vogelstein B, Albertsen H, White RL;
 PI Joslyn G;
 XX
 DR WPI; 1994-316233/39.
 DR N-PSDE; AAQ72297.
 XX
 PT New human adenomatous polyposis coli DNA encoding tumour repressor - and
 PT derived primers and probes for diagnosis, prognosis and treatment of
 PT cancer.
 XX
 PS Claim 3; Fig 7; 113pp; English.
 XX
 CC AAQ72297 is a cDNA isolated from the human adenomatous polyposis coli
 CC (APC) gene, it encodes the tumour repressors described in AAR63507 and
 CC AAR63508. Determination of alterations in APC or its expression products,
 CC can be used for the diagnosis and prognosis of cancer e.g. colorectal,
 CC lung and breast tumours; and for determining predisposition to certain

CC cancers such as familial adenomatous polyposis (FAP) and Gardner's
CC syndrome. The wild type APC gene (or a part of it) can be used
CC therapeutically to restore gene function, while primers and probes
CC derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to
CC detect mutations. Also APC proteins or analogues can be administered to
CC compensate for a defective gene, and epithelial cells, or transgenic
CC animals carrying a mutated APC allele are useful for detecting
CC therapeutic agents able to suppress tumorigenesis. (Updated on 25-MAR-
CC 2003 to correct PA field.)
XX
SQ Sequence 2860 AA;
Query Match 99.8%; Score 14551; DB 2; Length 2860;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2842; Conservative 0; Mismatches 0; Indels 18; Gaps 1;
QY 2 AAASYDQLLKQVAKWENSNLQCELEDSNHLTKLETSANMKVYLKOLQSGIDEANA 61
DB 1 AAASYDQLLKQVAKWENSNLQCELEDSNHLTKLETSANMKVYLKOLQSGIDEANA 60
QY 62 SSGQIDLLERLKEINLDS-----SNFFGVKLRKMSLRYSYSGREGSV 103
DB 61 SSGQIDLLERLKEINLDSNFFGVKLRKMSLRYSYSGREGSV 120
QY 104 SRSRGCSVPVPMGSPFRGFVNGSRESTGYLEBKERSILLADLKEEKEKDWYIAQLQ 163
DB 121 SRSRGCSVPVPMGSPFRGFVNGSRESTGYLEBKERSILLADLKEEKEKDWYIAQLQ 180
QY 164 NLTKRIDSLPTENFSLQTLTRQLEVEARQIVAMEEQLGTCDMEKQARRIARIQ 223
DB 181 NLTKRIDSLPTENFSLQTLTRQLEVEARQIVAMEEQLGTCDMEKQARRIARIQ 240
QY 224 IEKDILRIQLQSQATEARSSQKHETGSHDAERONEGQGVGEINWATSGNGQSGTTR 283
DB 241 IEKDILRIQLQSQATEARSSQKHETGSHDAERONEGQGVGEINWATSGNGQSGTTR 300
QY 284 MDHETASVSSSPHSPAPRLTSHLGTKEVYVLSLMLGTHDKDMSRTLLAMSSQDS 343
DB 301 MDHETASVSSSPHSPAPRLTSHLGTKEVYVLSLMLGTHDKDMSRTLLAMSSQDS 360
QY 344 CISMRQSGCLPLLQLLHGNDKDSVLLGNRSGSKEARARASALHNIHSPDDKRRRE 403
DB 361 CISMRQSGCLPLLQLLHGNDKDSVLLGNRSGSKEARARASALHNIHSPDDKRRRE 420
QY 404 IRVLHLEQIRAYCETCWEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHR 463
DB 421 IRVLHLEQIRAYCETCWEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHR 480
QY 464 HAMNELGLOAIABELLOVDCWEMGLTNDHYSITLRRVAGMALTNLTFGDVANKATLCMSK 523
DB 481 HAMNELGLOAIABELLOVDCWEMGLTNDHYSITLRRVAGMALTNLTFGDVANKATLCMSK 540
QY 524 GCMRALVAQLKSESEDLQOVVIASVLRNLSWADVNSKKTLEVEGSKVAMECALEVVKES 583
DB 541 GCMRALVAQLKSESEDLQOVVIASVLRNLSWADVNSKKTLEVEGSKVAMECALEVVKES 600
QY 584 TLKSVLSALWNLASHCTENKADI CAVDGALAFVGLTYRSQTNLTALIESGGGILRNVS 643
DB 601 TLKSVLSALWNLASHCTENKADI CAVDGALAFVGLTYRSQTNLTALIESGGGILRNVS 660
QY 644 SLIATNEDHROITLRENNCLQTLQHLKSHSLTIVSNACGTLWNLSARNPKDQALWDWGA 703
DB 661 SLIATNEDHROITLRENNCLQTLQHLKSHSLTIVSNACGTLWNLSARNPKDQALWDWGA 720
QY 704 VSMKXNLTHSKHKTAMGSAALRNLMANRPKYKDNIMPGSLPSLHVYKOKALEAE 763
DB 721 VSMKXNLTHSKHKTAMGSAALRNLMANRPKYKDNIMPGSLPSLHVYKOKALEAE 780
QY 764 LDAQHLSSETFDNIDNLSPKASHRSKQHKQSLYGDYVFTNRHDDNRDNFTNGMTVLS 823
DB 781 LDAQHLSSETFDNIDNLSPKASHRSKQHKQSLYGDYVFTNRHDDNRDNFTNGMTVLS 840
QY 824 PYLNTTVLPSSSSSRGSLDSSRSKDRSLERERIGLGNYPATENPOTSSKRGLOISTT 883

DB 841 PYLNTTVLPSSSSSRGSLDSSRSKDRSLERERIGLGNYPATENPOTSSKRGLOISTT 900
QY 884 AAQIAKWEVSAIHTSQEDSSGSTTELHCVTDERNALRRSSAAHTSHNTYNTKSENS 943
DB 901 AAQIAKWEVSAIHTSQEDSSGSTTELHCVTDERNALRRSSAAHTSHNTYNTKSENS 960
QY 944 NRTCSMPYAKLEYKRSNDLSNVSDNGYKRGOMKPSIESYSEDEDESKFCYQYPAD 1003
DB 961 NRTCSMPYAKLEYKRSNDLSNVSDNGYKRGOMKPSIESYSEDEDESKFCYQYPAD 1020
QY 1004 LAHKTHSANHMDNDGELDTPIYNSLYSDQLNSGRQSPQNERWARPKHIIIDEIKQS 1063
DB 1021 LAHKTHSANHMDNDGELDTPIYNSLYSDQLNSGRQSPQNERWARPKHIIIDEIKQS 1080
QY 1064 EQRQSRNQSTTVPVVTESTDDKHLKFQPHFQOQECVSPYRSGANGSETNRVSGNHGQ 1123
DB 1081 EQRQSRNQSTTVPVVTESTDDKHLKFQPHFQOQECVSPYRSGANGSETNRVSGNHGQ 1140
QY 1124 NVYSQLCQEDDYEDDKPTNYSERYSSEBQHEEERPTNYSIKYNEEKHVVQOPIDYSUKY 1183
DB 1141 NVYSQLCQEDDYEDDKPTNYSERYSSEBQHEEERPTNYSIKYNEEKHVVQOPIDYSUKY 1200
QY 1184 AYDIPISSQKQSFSSKSSGSSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRSGOP 1243
DB 1201 AYDIPISSQKQSFSSKSSGSSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRSGOP 1260
QY 1244 QAAATCKVSSINQETIOTYCYVEDTPICFRCSLSLSLSAEDICGNOTTOEADSNATLO 1303
DB 1261 QAAATCKVSSINQETIOTYCYVEDTPICFRCSLSLSLSAEDICGNOTTOEADSNATLO 1320
QY 1304 IABIKGKIGTRAEADPVSEPAVSOHPRTKSRLOGSSLSSESARHKAVERPFGSAKSPSK 1363
DB 1321 IABIKGKIGTRAEADPVSEPAVSOHPRTKSRLOGSSLSSESARHKAVERPFGSAKSPSK 1380
QY 1364 SGAQTPKSPPEHYVQETPLMFSCRCTSVSSLSQSFESRSIASSVQSPGCMVSGIISPSDL 1423
DB 1381 SGAQTPKSPPEHYVQETPLMFSCRCTSVSSLSQSFESRSIASSVQSPGCMVSGIISPSDL 1440
QY 1424 PDSPGOTMPPSKSKTPPPPPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLP 1483
DB 1441 PDSPGOTMPPSKSKTPPPPPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLP 1500
QY 1484 DADTLLHFATETPPGDFSCSSLSALSDELPEPIQKDELIRIMPPVQENDNGNETSEQPK 1543
DB 1501 DADTLLHFATETPPGDFSCSSLSALSDELPEPIQKDELIRIMPPVQENDNGNETSEQPK 1560
QY 1544 ESNENOKEAEKTIIDSEKDLDDDDDDIIELEECIIISAMPTKSRKGGKPKQATSKLPP 1603
DB 1561 ESNENOKEAEKTIIDSEKDLDDDDDDIIELEECIIISAMPTKSRKGGKPKQATSKLPP 1620
QY 1604 PVARKEPSQLPVYKLLPSQNRLOPQKHVSFTPGDMPRVYCVBGTPIINFSTATSLSLTIE 1663
DB 1621 PVARKEPSQLPVYKLLPSQNRLOPQKHVSFTPGDMPRVYCVBGTPIINFSTATSLSLTIE 1680
QY 1664 SPPELAAAGEVVRGGAQSGEFKRDITPTEGRSTDBAQGGKTSVVTIPELDDNKABEGDI 1723
DB 1681 SPPELAAAGEVVRGGAQSGEFKRDITPTEGRSTDBAQGGKTSVVTIPELDDNKABEGDI 1740
QY 1724 LAECINSAMPKSKHHPFRVKIMDOVQAASASSAPNKNQDQKGGKKTSPVKPIPQNT 1783
DB 1741 LAECINSAMPKSKHHPFRVKIMDOVQAASASSAPNKNQDQKGGKKTSPVKPIPQNT 1800
QY 1784 EYRTRVRKXNADSKXNLNAERVSFNDKSKQNLKNNKDFNDKLPNNEDVRGSGFAFDP 1843
DB 1801 EYRTRVRKXNADSKXNLNAERVSFNDKSKQNLKNNKDFNDKLPNNEDVRGSGFAFDP 1860
QY 1844 HHVTPTEGTPYCFSRNDSLSLDFDDDDVDLSREKALRKAKENKESAKVTSHTLTSN 1903
DB 1861 HHVTPTEGTPYCFSRNDSLSLDFDDDDVDLSREKALRKAKENKESAKVTSHTLTSN 1920
QY 1904 QQSANKTQALAKOPINRGOPKPILOKQSTFPQSSKIDPRGAATDEKLOJFAJENTPVCF 1963

Db 1921 QOSANKTQAIKQPINRGQPKPILOKOSTFPQSSKDIPDRGAATDBKLNFAIENTPVCF 1980
 Qy 1964 SHNSLSLSLDIDQENNNKNEPIKETEPDPSQEPSPKQASGYAPKSHFVEDTPVCF 2023
 Db 1981 SHNSLSLSLDIDQENNNKNEPIKETEPDPSQEPSPKQASGYAPKSHFVEDTPVCF 2040
 Qy 2024 NSLSLSLSIDSEDDLLQECISSAMPKKPSRLKGNEXHSPNNMGIGLDCITLCLKDI 2083
 Db 2041 NSLSLSLSIDSEDDLLQECISSAMPKKPSRLKGNEXHSPNNMGIGLDCITLCLKDI 2100
 Qy 2084 QRPDEHGLSPDENFDWKAIOGANSIVSSLHQAACLSQAASDSDSILSLKSGIS 2143
 Db 2101 QRPDEHGLSPDENFDWKAIOGANSIVSSLHQAACLSQAASDSDSILSLKSGIS 2160
 Qy 2144 LGPPFHLTPDOEKPTNKGPRILKPGKSTLTETKIESEKGIKGGKVKVKSILITGV 2203
 Db 2161 LGPPFHLTPDOEKPTNKGPRILKPGKSTLTETKIESEKGIKGGKVKVKSILITGV 2220
 Qy 2204 RSNSEISGOMKQPLQANMPSISRGRTMIHPIGVNNSSTSPSVKGPPLKTPAKSPSE 2263
 Db 2221 RSNSEISGOMKQPLQANMPSISRGRTMIHPIGVNNSSTSPSVKGPPLKTPAKSPSE 2280
 Qy 2264 GQTATTSRGAKPSVKSSELPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQS 2323
 Db 2281 GQTATTSRGAKPSVKSSELPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQS 2340
 Qy 2324 PGRNSISPGNGISPPNKLSQLPRTSPSTASTKSSGSGKMSYTPSGROMSQOQNTKQTG 2383
 Db 2341 PGRNSISPGNGISPPNKLSQLPRTSPSTASTKSSGSGKMSYTPSGROMSQOQNTKQTG 2400
 Qy 2384 LSKNASSIPRESASKGLNOMNGANKVLSRMSSTKSSGSESDRSERPVLRQSTF 2443
 Db 2401 LSKNASSIPRESASKGLNOMNGANKVLSRMSSTKSSGSESDRSERPVLRQSTF 2460
 Qy 2444 IKAPSPTRLKLEASPESESIPSRPASPTRSQAQTPVLSPLPDMSLTSHSVQAGG 2503
 Db 2461 IKAPSPTRLKLEASPESESIPSRPASPTRSQAQTPVLSPLPDMSLTSHSVQAGG 2520
 Qy 2504 WRKLPLNLSTIYDGRPAKHDIAHSHSESPSLPINRSGTWKREKSHSSSLPRVST 2563
 Db 2521 WRKLPLNLSTIYDGRPAKHDIAHSHSESPSLPINRSGTWKREKSHSSSLPRVST 2580
 Qy 2564 WRTGSSSILSASSSESEKAKSEDEKHVNSISGKQKQENQVSAKGTWRKIKENEFSP 2623
 Db 2581 WRTGSSSILSASSSESEKAKSEDEKHVNSISGKQKQENQVSAKGTWRKIKENEFSP 2640
 Qy 2624 NSTSQTVSSGATNGABSKTLIYQMAPAVSKTEDVWVREDCPINNPRGRSPGTGTPV 2683
 Db 2641 NSTSQTVSSGATNGABSKTLIYQMAPAVSKTEDVWVREDCPINNPRGRSPGTGTPV 2700
 Qy 2684 DSVSEKANPNIKDKONQAKQNGVSGVPMRTVGLNRLTFSIQVADPDQKGTETKPGQ 2743
 Db 2701 DSVSEKANPNIKDKONQAKQNGVSGVPMRTVGLNRLTFSIQVADPDQKGTETKPGQ 2760
 Qy 2744 NPVPVSETHESIVERTPSSSSSKHSPSPGTVAARTVPPNPNPSPKSGADSTSAAPS 2803
 Db 2761 NPVPVSETHESIVERTPSSSSSKHSPSPGTVAARTVPPNPNPSPKSGADSTSAAPS 2820
 Qy 2804 QIPTPVNNNTKRDSTKDTSESSGTSQPKRHSGSVLVTSV 2843
 Db 2821 QIPTPVNNNTKRDSTKDTSESSGTSQPKRHSGSVLVTSV 2860

RESULT 6
 AAW76140
 ID AAW76140 standard; protein; 2843 AA.
 XX
 AC AAW76140;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-NOV-1998 (first entry)
 XX
 DE Human APC protein #1.

XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
 KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
 KW Gardner's Syndrome; GS; predisposition.
 XX Homo sapiens.
 XX US5783666-A.
 XX 21-JUL-1998.
 XX 25-MAY-1995; 95US-00452655.
 XX 16-JAN-1991; 91GB-00000962.
 XX 16-JAN-1991; 91GB-00000963.
 XX 16-JAN-1991; 91GB-00000974.
 XX 16-JAN-1991; 91GB-00000975.
 XX 08-AUG-1991; 91US-007411940.
 XX 12-AUG-1994; 94US-00289548.
 XX (CANC-) CANCER INST.
 PA (UNJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 PA (ZENE) ZENECA PHARM.
 XX Kinzler K, Joslyn G, Markham AF, Carlson M, White RL;
 PI Thiliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;
 PI Vogelstein B, Hedge P;
 XX WPI: 1998-427100/36.
 DR N-PSDB; AAV56447.
 XX Adenomatous polyposis coli protein - useful in the treatment of cancers
 PT associated with mutation(s) on human chromosome 5q21.
 XX Disclosure; Col 41-54; 102pp; English.

XX This sequence represents a human familial adenomatous polyposis coli
 CC (APC) protein from clone DP2.5. The gene for the protein is present on
 CC human chromosome 5q21 and is also referred to as adenomatous polyposis
 CC coli gene. It is a tumour suppressor gene, and mutations in this gene
 CC have been associated with tumourigenesis in retinoblastoma and colorectal
 CC tumours, and especially familial adenomatous polyposis (FAP) and
 CC Gardner's Syndrome (GS). The protein can be used in therapy to replace
 CC lack of native functional protein and the nucleic acids can be used for
 CC gene therapy. The nucleic acids that encode them can also be used as
 CC probes and primers in detection of the cancers and predisposition to it.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX Sequence 2843 AA;

Query Match 99.7%; Score 14533; DB 2; Length 2843;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAAASYDQLLKQVEALKMENSLNRQLEDNSNHLTKLETEASNKKEVLKQLOQSIEDAM 60
 Db 1 MAAASYDQLLKQVEALKMENSLNRQLEDNSNHLTKLETEASNKKEVLKQLOQSIEDAM 60
 Qy 61 ASSGQIDLLERLKEINLDSNPFVKLRSMKSLRSYSGREGSVSRGSCSPVPMGSPFR 120
 Db 61 ASSGQIDLLERLKEINLDSNPFVKLRSMKSLRSYSGREGSVSRGSCSPVPMGSPFR 120
 Qy 121 RGFVNGSRSTGYLEELERKERSILLADLDKEKEKDWYVAQQLNLTTRIDSLPTENFSL 180
 Db 121 RGFVNGSRSTGYLEELERKERSILLADLDKEKEKDWYVAQQLNLTTRIDSLPTENFSL 180
 Qy 181 QTDLTRQLEYEARQIRVANBEQOLGTCQDMEKRAQRARIQOIEKDLIRIQLQSQAT 240
 Db 181 QTDLTRQLEYEARQIRVANBEQOLGTCQDMEKRAQRARIQOIEKDLIRIQLQSQAT 240
 Qy 241 EASRSQKHETGSHDAERQNEGQGVGEINMATSNGOGSTTRMDHETASVLSSTSTSA 300

Db 241 EAERSQNKHETGSHDAERQEGQVGEINMATSGNGOGSTTRMDHETASVLSSTSTSA 300
Qy 301 PRLTSLHGTKEVMPYVLLSMLGTHDKDDMDRTLLAMSSODSCISWRQSCCLPLLQILL 360
Db 301 PRLTSLHGTKEVMPYVLLSMLGTHDKDDMDRTLLAMSSQDSCISWRQSCCLPLLQILL 360
Qy 361 HGNDKDSVLLGNRSGSKARAPASAALENIIHSQPDCKRGRREIRVLLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNRSGSKARAPASAALENIIHSQPDCKRGRREIRVLLHLEQIRAYCETC 420
Qy 421 WQQAHAPEGMQDQKNPMPAPVEHOICPAVCVLMKLSFDEBHRHAMNELLGQIAIELLO 480
Db 421 WQQAHAPEGMQDQKNPMPAPVEHOICPAVCVLMKLSFDEBHRHAMNELLGQIAIELLO 480
Qy 481 VDCMYGLTNDHYSITLRYAGMALTNLTFFGVANKATLCSMKGCMRALVAQLKSESEDL 540
Db 481 VDCMYGLTNDHYSITLRYAGMALTNLTFFGVANKATLCSMKGCMRALVAQLKSESEDL 540
Qy 541 QOVIASVLRNLNWRADVNSKTLTRVGSVKALMECALEVKKESTLKVLSALMNLNLSAHT 600
Db 541 QOVIASVLRNLNWRADVNSKTLTRVGSVKALMECALEVKKESTLKVLSALMNLNLSAHT 600
Qy 601 ENKADICAVDGAFLVGLTVYRSQNTNLATIESGGGILRVNVSLLIATNEDHROILRENN 660
Db 601 ENKADICAVDGAFLVGLTVYRSQNTNLATIESGGGILRVNVSLLIATNEDHROILRENN 660
Qy 661 CLQTLLOHLKSHSLTVSNACOTLNLGARNPKQOALWDMGVMSMLNLIHSHKHMIAM 720
Db 661 CLQTLLOHLKSHSLTVSNACOTLNLGARNPKQOALWDMGVMSMLNLIHSHKHMIAM 720
Qy 721 GSAALRNLMANRPKAKYKXANIMSPGSSLPVLHVRKQKALEAEALDAQHLSETPDNIOLNLS 780
Db 721 GSAALRNLMANRPKAKYKXANIMSPGSSLPVLHVRKQKALEAEALDAQHLSETPDNIOLNLS 780
Qy 781 PKASHRSKORHQSILYGYVFTDNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSSRGS 840
Db 781 PKASHRSKORHQSILYGYVFTDNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSSRGS 840
Qy 841 LBSRSSEKDRSLERERGLGNHYHPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900
Db 841 LBSRSSEKDRSLERERGLGNHYHPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900
Qy 901 QEDRSSGSTTEHLVCTDERNALRRSSAAHTSNFTYNTFKSENSNRCTCSPYAKLEYKRSS 960
Db 901 QEDRSSGSTTEHLVCTDERNALRRSSAAHTSNFTYNTFKSENSNRCTCSPYAKLEYKRSS 960
Qy 961 NDLNSVSSNDGYKRGQWKPSIESYSEDDSKFCYQYPADLAHAKTHSANHMDNDGE 1020
Db 961 NDLNSVSSNDGYKRGQWKPSIESYSEDDSKFCYQYPADLAHAKTHSANHMDNDGE 1020
Qy 1021 LPTPINYSKYDEQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Db 1021 LPTPINYSKYDEQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFPQHFQOECVPSVRSRGANGSETNRVGSNHGINQVNSQSLCOEDDIEDDKP 1140
Db 1081 STDDKHLKFPQHFQOECVPSVRSRGANGSETNRVGSNHGINQVNSQSLCOEDDIEDDKP 1140
Qy 1141 TNYSERYSEREEHBEERTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
Db 1141 TNYSERYSEREEHBEERTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSEKS 1200
Qy 1201 SSCQSKTTHMSSSENTSTPSSNAKRONOLHPSAOSRSCOPAKATCKVSSINQSTIQ 1260
Db 1201 SSCQSKTTHMSSSENTSTPSSNAKRONOLHPSAOSRSCOPAKATCKVSSINQSTIQ 1260
Qy 1261 TYCVEDTPICFGRCSLSLSAEDEIGCNQTTQEADSAANTLQIAEIKGKITRSADDPV 1320
Db 1261 TYCVEDTPICFGRCSLSLSAEDEIGCNQTTQEADSAANTLQIAEIKGKITRSADDPV 1320
Qy 1321 SEVPASVOHPRTKSRLOGSSLSLSARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVOET 1380
Db 1321 SEVPASVOHPRTKSRLOGSSLSLSARHKAVEFPSSGAKSPKSGAQTPKSPPEHYVOET 1380

Qy 1381 PLMFRCCTSVSLSLDSFBRSSTASSVQSEPCSGMWVSGIISPSDLDPSPCQOTMPPSRSKTTP 1440
Db 1381 PLMFRCCTSVSLSLDSFBRSSTASSVQSEPCSGMWVSGIISPSDLDPSPCQOTMPPSRSKTTP 1440
Qy 1441 PFPQTAQTKREVPKNKAFTAEKRESGPKQAANAQVQVLPADATLLHLPATESTPDGF 1500
Db 1441 PFPQTAQTKREVPKNKAFTAEKRESGPKQAANAQVQVLPADATLLHLPATESTPDGF 1500
Qy 1501 SCSSLSLSALSDPEPIQKDVLRIMPVQVNDNGNETESBQPKESNENQEKAEKTTDSE 1560
Db 1501 SCSSLSLSALSDPEPIQKDVLRIMPVQVNDNGNETESBQPKESNENQEKAEKTTDSE 1560
Qy 1561 KDLDDSDDDDIIELEECIIISAMPTKSRKQKPAQATSKLPPPVARKPSQOLPVYKLLPS 1620
Db 1561 KDLDDSDDDDIIELEECIIISAMPTKSRKQKPAQATSKLPPPVARKPSQOLPVYKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATSLDLTIESPPNELAAGEVRGGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATSLDLTIESPPNELAAGEVRGGAQ 1680
Qy 1681 SEFEKROTITPEGRSTDEAOGGKTSVVTIPELDDNKAEEGDILAEICINSAMPKKGSHKP 1740
Db 1681 SEFEKROTITPEGRSTDEAOGGKTSVVTIPELDDNKAEEGDILAEICINSAMPKKGSHKP 1740
Qy 1741 FRVKIMQVOQASASSAPNKNQIDGKKKPTSPVVKPIPONTBYRTRVRKNADSKNMLN 1800
Db 1741 FRVKIMQVOQASASSAPNKNQIDGKKKPTSPVVKPIPONTBYRTRVRKNADSKNMLN 1800
Qy 1801 ARVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVGRGFAPDSPHHYTPIEGTTCFCSRND 1860
Db 1801 ARVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVGRGFAPDSPHHYTPIEGTTCFCSRND 1860
Qy 1861 SLSSLDLDFDDDDVLSREKAEIRKAKENKESAKVTSTELTNSQOSANKTOAIKQPINR 1920
Db 1861 SLSSLDLDFDDDDVLSREKAEIRKAKENKESAKVTSTELTNSQOSANKTOAIKQPINR 1920
Qy 1921 GQPKPILQKQSTFPQSSKDIIDRGAATDEKLQNFALIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILQKQSTFPQSSKDIIDRGAATDEKLQNFALIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 NKNEPFIKETBPPDSQGPSPQASGVAPKGFHVEDTPVCFPSRNSISSLSIDSEDLQ 2040
Db 1981 NKNEPFIKETBPPDSQGPSPQASGVAPKGFHVEDTPVCFPSRNSISSLSIDSEDLQ 2040
Qy 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGGLGDELTLDLKDIOQRPDSEHGLSPDSNFD 2100
Db 2041 ECISAMPKPKKPSRLKGDNEKHSPRNMGGLGDELTLDLKDIOQRPDSEHGLSPDSNFD 2100
Qy 2101 WFAIPEGANSIVSSLHQAAAAACLSRQASSDSDSILSKSGISILGSPFHLTPDQEEKPFT 2160
Db 2101 WFAIPEGANSIVSSLHQAAAAACLSRQASSDSDSILSKSGISILGSPFHLTPDQEEKPFT 2160
Qy 2161 SNKGPRILKPEKSTLETKTIESKGIKGGKVVYKSLITGKVRNSGEISQMKQIOLAN 2220
Db 2161 SNKGPRILKPEKSTLETKTIESKGIKGGKVVYKSLITGKVRNSGEISQMKQIOLAN 2220
Qy 2221 MPISIRGRMTIHIPVRNNSSTSPVSKGPPKLTTPASKSPSEGTATTSPRGAQPVKS 2280
Db 2221 MPISIRGRMTIHIPVRNNSSTSPVSKGPPKLTTPASKSPSEGTATTSPRGAQPVKS 2280
Qy 2281 ELSPVARTSQIGSSKAPSRSGSRDSTPSRPAQQLSRPTQSPGRNSISPRNGISPPN 2340
Db 2281 ELSPVARTSQIGSSKAPSRSGSRDSTPSRPAQQLSRPTQSPGRNSISPRNGISPPN 2340
Qy 2341 KLSQLPRTSSPSTASTKSGGSKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASK 2400
Db 2341 KLSQLPRTSSPSTASTKSGGSKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASK 2400
Qy 2401 LNMKNNGANKKVELSRMSTKSGSSEDRSERPVLVROSTFIKEAPSPFLRKLBSA 2460
Db 2401 LNMKNNGANKKVELSRMSTKSGSSEDRSERPVLVROSTFIKEAPSPFLRKLBSA 2460

QY 2461 SFESLSPSSRPASPTRSOATFVLSPSLPDMSLSTSHSVQAGGWRKLPNLSPTIYNDG 2520
 DB |||||
 QY 2461 SFESLSPSSRPASPTRSOATFVLSPSLPDMSLSTSHSVQAGGWRKLPNLSPTIYNDG 2520
 DB |||||
 QY 2521 RPAKRHHIARSHSPRLPINSRGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 DB |||||
 QY 2521 RPAKRHHIARSHSPRLPINSRGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 DB |||||
 QY 2581 SEKAKSDEKHVNSISGTSKQENQVSAKGTWKIKENEFSPNISTQTVSSGATNGAES 2640
 DB |||||
 QY 2581 SEKAKSDEKHVNSISGTSKQENQVSAKGTWKIKENEFSPNISTQTVSSGATNGAES 2640
 DB |||||
 QY 2641 KTLIIQWAPAVSKTEDVWVRIEDCPINNPSPRGSPGTGNTPPVIDSVSEKAPNPKDSKDN 2700
 DB |||||
 QY 2641 KTLIIQWAPAVSKTEDVWVRIEDCPINNPSPRGSPGTGNTPPVIDSVSEKAPNPKDSKDN 2700
 DB |||||
 QY 2701 QAKQNVGSGVPWRTVGLNRLTSFTQVDAPDQKGTETIKPGQNNPVPVSETNESPIVERT 2760
 DB |||||
 QY 2701 QAKQNVGSGVPWRTVGLNRLTSFTQVDAPDQKGTETIKPGQNNPVPVSETNESPIVERT 2760
 DB |||||
 QY 2761 PFSSSSSSKHSFSGTVAARVTFPNYNPSPRKSSADSTGAPSOIPTPVNNNTKRDSDKT 2820
 DB |||||
 QY 2761 PFSSSSSSKHSFSGTVAARVTFPNYNPSPRKSSADSTGAPSOIPTPVNNNTKRDSDKT 2820
 DB |||||
 QY 2821 DSTEESGTSQPKHSGSYLTVS 2843
 DB |||||
 QY 2821 DSTEESGTSQPKHSGSYLTVS 2843
 DB |||||

RESULT 7
 AAW76144
 ID AAW76144 standard; protein; 2843 AA.
 AC AAW76144;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-NOV-1998 (first entry)
 XX
 DE Human APC protein #2.
 XX
 KW Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
 KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
 KW Gardner's Syndrome; GS; predisposition.
 XX
 OS Homo sapiens.
 XX
 PN US5783666-A.
 XX
 PD 21-JUL-1998.
 XX
 PF 25-MAY-1995; 95US-00452655.
 XX
 PR 16-JAN-1991; 91GB-00000962.
 PR 16-JAN-1991; 91GB-00000963.
 PR 16-JAN-1991; 91GB-00000974.
 PR 16-JAN-1991; 91GB-00000975.
 PR 08-AUG-1991; 91US-00741940.
 PR 12-AUG-1994; 94US-00289548.
 XX
 PA (CANC-) CANCER INST.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 PA (ZENE) ZENECA PHARM.
 XX
 PI Kinzler X, Joslyn G, Markham AF, Carlson M, White RL;
 PI Thliveris A, Groden J, Nakamura Y, Albertsen H;
 PI Vogelstein B, Hedge PJ;
 XX
 DR WPI; 1998-427100/36.
 XX
 PT Adenomatous polyposis coli protein - useful in the treatment of cancers
 PT associated with mutation(s) on human chromosome 5q21.
 XX

PS Disclosure; Col 63-78; 102pp; English.
 XX
 CC This sequence represents a human familial adenomatous polyposis coli
 CC (APC) protein isolated from 87 cDNA clones. The gene for the protein is
 CC present on human chromosome 5q21 and is also referred to as adenomatous
 CC polyposis coli gene. It is a tumour suppressor gene, and mutations in
 CC this gene have been associated with tumorigenesis in retinoblastoma and
 CC colorectal tumours, and especially familial adenomatous polyposis (FAP)
 CC and Gardner's Syndrome (GS). The protein can be used in therapy to
 CC replace lack of native functional protein and the nucleic acids can be
 CC used for gene therapy. The nucleic acids that encode them can also be
 CC used as probes and primers in detection of the cancers and predisposition
 CC to it. (Updated on 25-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 2843 AA;

Query Match 99.7%; Score 14533; DB 2; Length 2843;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAASYPDQLLKQVEALQKXNSLRQLEEDNSNHLTKLETEASNKKEVLKQLOGSIEDEAM 60
 DB |||||
 QY 1 MAASYPDQLLKQVEALQKXNSLRQLEEDNSNHLTKLETEASNKKEVLKQLOGSIEDEAM 60
 DB |||||
 QY 61 ASSQOIDLRLKELNLDSSNFPVGLRSMVLSRYSYSGREGSVSSSGECSPVPMGSFPR 120
 DB |||||
 QY 61 ASSQOIDLRLKELNLDSSNFPVGLRSMVLSRYSYSGREGSVSSSGECSPVPMGSFPR 120
 DB |||||
 QY 121 RGFVNGSRESTGYLEEKERSLLDLADLKEEKEKDWYLAQLQNLTKRIDSPLTENFSL 180
 DB |||||
 QY 121 RGFVNGSRESTGYLEEKERSLLDLADLKEEKEKDWYLAQLQNLTKRIDSPLTENFSL 180
 DB |||||
 QY 181 QDLTRQLEVEARQIRVAMEEOLGTCQDMKRAQRIARIQOIEKDIIRIROLLOSQT 240
 DB |||||
 QY 181 QDLTRQLEVEARQIRVAMEEOLGTCQDMKRAQRIARIQOIEKDIIRIROLLOSQT 240
 DB |||||
 QY 241 EBERSQNHETGSHDAERQNEQGVGEINMATSNGQGGSTTRMDHETASVLSSTHSA 300
 DB |||||
 QY 241 EBERSQNHETGSHDAERQNEQGVGEINMATSNGQGGSTTRMDHETASVLSSTHSA 300
 DB |||||
 QY 301 PRLTSLHGLTKVEMVYSLISMLGTHDKDMSRTLLAMSSQSCISMROSQGLPLLIQLL 360
 DB |||||
 QY 301 PRLTSLHGLTKVEMVYSLISMLGTHDKDMSRTLLAMSSQSCISMROSQGLPLLIQLL 360
 DB |||||
 QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHHSQPDCKEGRREIRVHLLEQIRAYCETC 420
 DB |||||
 QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHHSQPDCKEGRREIRVHLLEQIRAYCETC 420
 DB |||||
 QY 421 WEMQEAHEPGMDODKNMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOA-AELIQ 480
 DB |||||
 QY 421 WEMQEAHEPGMDODKNMPAPVEHQICPAVCLMKLSFDEEHRHAMNELGGLOA-AELIQ 480
 DB |||||
 QY 481 VDCMYGLTNDHYSITLRRYAGMALTNLTFFGVANKATLCSMKGCWRALVAOLKSESED 540
 DB |||||
 QY 481 VDCMYGLTNDHYSITLRRYAGMALTNLTFFGVANKATLCSMKGCWRALVAOLKSESED 540
 DB |||||
 QY 541 QQVIASVNLNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALWNLNLSAHT 600
 DB |||||
 QY 541 QQVIASVNLNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALWNLNLSAHT 600
 DB |||||
 QY 601 ENKADICAVDGLAFVLVGLTYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 660
 DB |||||
 QY 601 ENKADICAVDGLAFVLVGLTYRSQNTLAIIESGGILRNVSLLIATNEDHRQLRENN 660
 DB |||||
 QY 661 CLQTLQLKSHSLTIVSNACOTLWNLARNPKDQALWDMGAVSMKLNLSHSHKMTAM 720
 DB |||||
 QY 661 CLQTLQLKSHSLTIVSNACOTLWNLARNPKDQALWDMGAVSMKLNLSHSHKMTAM 720
 DB |||||
 QY 721 GSAAALRNLMANRPKYKDANIMSPGSSLPVHVKOKALEAELDAQHLSHSHKMTAM 780
 DB |||||
 QY 721 GSAAALRNLMANRPKYKDANIMSPGSSLPVHVKOKALEAELDAQHLSHSHKMTAM 780
 DB |||||

Db 781 PKASHRSQRKHQSLGYDYVFTDTHRDDNRSDNFTGNMVTLSPLYLNTTVLPSSSSSRGS 840
Qy 841 LDSRSEKDRSLERERGIQGLNHYHPATENPGTSSKRGLOIQTSTAQAIAKVAEEVSAIHTS 900
Db 841 LDSRSEKDRSLERERGIQGLNHYHPATENPGTSSKRGLOIQTSTAQAIAKVAEEVSAIHTS 900
Qy 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNTYNTFKSENSNRTCSMPYAKULEYKRSS 960
Db 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNTYNTFKSENSNRTCSMPYAKULEYKRSS 960
Qy 961 NDSLNSVSSNDGYKRGOMKPSIESYSDDSKFCVSYQVADLAHKITHSANHMDNDGE 1020
Db 961 NDSLNSVSSNDGYKRGOMKPSIESYSDDSKFCVSYQVADLAHKITHSANHMDNDGE 1020
Qy 1021 LDTPIVNSLYKSDQLNSGRSPQSNERWARPKHI IEDEIKQSEQRQSNQSTTPYVTE 1080
Db 1021 LDTPIVNSLYKSDQLNSGRSPQSNERWARPKHI IEDEIKQSEQRQSNQSTTPYVTE 1080
Qy 1081 STDDKHLKFPQHFQOQECVSPYRERGANGSTNRVGNHGINQVNSQSLCOEDDYEDDKP 1140
Db 1081 STDDKHLKFPQHFQOQECVSPYRERGANGSTNRVGNHGINQVNSQSLCOEDDYEDDKP 1140
Qy 1141 TNSERYSEEEQHBEERPTNYSIKYNEKRVDPQIDYSLKYATDIPSSQKQSFSPKS 1200
Db 1141 TNSERYSEEEQHBEERPTNYSIKYNEKRVDPQIDYSLKYATDIPSSQKQSFSPKS 1200
Qy 1201 SSGOSSKTEHSSSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Db 1201 SSGOSSKTEHSSSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFRSCSSLSLSAASDEIGCNQTTQEADSANTLOIAEIKKIGITSABDPV 1320
Db 1261 TYCVEDTPICFRSCSSLSLSAASDEIGCNQTTQEADSANTLOIAEIKKIGITSABDPV 1320
Qy 1321 SEVPAVSQHPRTKSLRLOGSLSSSARHKAVEFSPGAKSPSKGAQTPKSPPHYVOET 1380
Db 1321 SEVPAVSQHPRTKSLRLOGSLSSSARHKAVEFSPGAKSPSKGAQTPKSPPHYVOET 1380
Qy 1381 PLMFRCSTSVSLDSFEBSRTASSVQSEPCSGMWGIIISPSDLDPSPGQTPPERSKTTP 1440
Db 1381 PLMFRCSTSVSLDSFEBSRTASSVQSEPCSGMWGIIISPSDLDPSPGQTPPERSKTTP 1440
Qy 1441 PPPOTAQTKREVPKNKAPTAKRSGPKQAANAQVQVLPDADTLHPFATESTPDGF 1500
Db 1441 PPPOTAQTKREVPKNKAPTAKRSGPKQAANAQVQVLPDADTLHPFATESTPDGF 1500
Qy 1501 SCSSLSALSLEDEPIQKDVLRIMPVQENDNGNETESEOPKESNENQKEAEKTIIDSE 1560
Db 1501 SCSSLSALSLEDEPIQKDVLRIMPVQENDNGNETESEOPKESNENQKEAEKTIIDSE 1560
Qy 1561 KDLLDSDSDDDIEILEECIIISAMPTKSSRKKGKPAQATASKLPPPVARKPSOLPYVKLLPS 1620
Db 1561 KDLLDSDSDDDIEILEECIIISAMPTKSSRKKGKPAQATASKLPPPVARKPSOLPYVKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDDMPRVYCEGTPIINSTATSISDLTIESPPNELAAGEVGRGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDDMPRVYCEGTPIINSTATSISDLTIESPPNELAAGEVGRGAQ 1680
Qy 1681 SGEFEKRTIITEGRSTDEAOGKTSSTVI IPELDNKAEEGDIIAECINSAMPKKGSHK 1740
Db 1681 SGEFEKRTIITEGRSTDEAOGKTSSTVI IPELDNKAEEGDIIAECINSAMPKKGSHK 1740
Qy 1741 FRVKKIMDVOQOASASSAPKNNQLDGKKKTPSPVKPI PONTYRTRVRKNADSKNNLN 1800
Db 1741 FRVKKIMDVOQOASASSAPKNNQLDGKKKTPSPVKPI PONTYRTRVRKNADSKNNLN 1800
Qy 1801 AERVFSDNKDSKKQNLKNNKDFNDKLPNNEDRVGSPFAPSPHHYPIEGTPYCFSRND 1860
Db 1801 AERVFSDNKDSKKQNLKNNKDFNDKLPNNEDRVGSPFAPSPHHYPIEGTPYCFSRND 1860
Qy 1861 SLSSLDFFDDVDLSREKAELRAKENKESBAKVTSHLTSTNQOSANKTOIAIAKQPINR 1920

Db 1861 SLSSLDFFDDVDLSREKAELRAKENKESBAKVTSHLTSTNQOSANKTOIAIAKQPINR 1920
Qy 1921 GQPKPILOKQSTFPQSSKDI PDGGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILOKQSTFPQSSKDI PDGGAATDEKLNQFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 NKENEDI KETETEDDSCGEBSPKQASGYAPKSPHVEDTPVCFSNSLSLSLSIDSEDDLLQ 2040
Db 1981 NKENEDI KETETEDDSCGEBSPKQASGYAPKSPHVEDTPVCFSNSLSLSLSIDSEDDLLQ 2040
Qy 2041 ECITSSAMPKKKPSRLKGDNEKHSRNMGGIIGEDITLDLKDIQRPDSEHGLSPQSENFD 2100
Db 2041 ECITSSAMPKKKPSRLKGDNEKHSRNMGGIIGEDITLDLKDIQRPDSEHGLSPQSENFD 2100
Qy 2101 WKAIQEGANSIVSSLHQAAAAACLSQASDSDSDLSLSGSLSGSPFHLTPDQBEKPT 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACLSQASDSDSDLSLSGSLSGSPFHLTPDQBEKPT 2160
Qy 2161 SNKGPRILAPGKSTLTETTKIESESKGIGKKVKYKSLITGVRNSSEISGOMKQPLQAN 2220
Db 2161 SNKGPRILAPGKSTLTETTKIESESKGIGKKVKYKSLITGVRNSSEISGOMKQPLQAN 2220
Qy 2221 MFSISGRRTMIHIPGVNRSSSTSPVSKGPPKLPASKSPSEGGQTATTSPRGAKPVS 2280
Db 2221 MFSISGRRTMIHIPGVNRSSSTSPVSKGPPKLPASKSPSEGGQTATTSPRGAKPVS 2280
Qy 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQOPLSRPIOSPGRNSISPGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGGSSKAPSRSGRSDTSPRPAQOPLSRPIOSPGRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSPSTASTKSSGSGKMYTSPGRQMSQOQLTKOTGLSKNASSIPRSESASK 2400
Db 2341 KLSQLPRTSPSTASTKSSGSGKMYTSPGRQMSQOQLTKOTGLSKNASSIPRSESASK 2400
Qy 2401 LNOMNNGANKVIELSRMSSYKSSGESDRSERPVLVROSTFIKEAPSPTRLRKLSEA 2460
Db 2401 LNOMNNGANKVIELSRMSSYKSSGESDRSERPVLVROSTFIKEAPSPTRLRKLSEA 2460
Qy 2461 SFESLSPPSRPASPTRSQAQTPVLSPLDMSLSTHSSVQAGGWRLPPNLSPITIEYNDG 2520
Db 2461 SFESLSPPSRPASPTRSQAQTPVLSPLDMSLSTHSSVQAGGWRLPPNLSPITIEYNDG 2520
Qy 2521 RPAKRHDIAARSHSESPRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSES 2580
Db 2521 RPAKRHDIAARSHSESPRLPINRSGTWKREHSHSSLPVSTWRTTSGSSSILSASSES 2580
Qy 2581 SEKAKSEDEKHVNSISGTTKQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGTTKQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
Qy 2641 KTLIYQMAPAVKTEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKANPNIKDSKDN 2700
Db 2641 KTLIYQMAPAVKTEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKANPNIKDSKDN 2700
Qy 2701 QAKQNYGNGSVPMRTVGLNRLTSFTQVDAPOKGTETIKPGQNNPVVSETNESPIVERT 2760
Db 2701 QAKQNYGNGSVPMRTVGLNRLTSFTQVDAPOKGTETIKPGQNNPVVSETNESPIVERT 2760
Qy 2761 PFSSSSSSKHSPPSGTVAARVTPFNTNPSFRKSSADSTARSQIPPTPVNNTTKEDSKT 2820
Db 2761 PFSSSSSSKHSPPSGTVAARVTPFNTNPSFRKSSADSTARSQIPPTPVNNTTKEDSKT 2820
Qy 2821 DSTESSGTQSPKHSGLYVTSV 2843
Db 2821 DSTESSGTQSPKHSGLYVTSV 2843

RESULT 8
AAB23011
ID AAB23011 standard; protein; 2843 AA.
XX
XX AAB23011;
XX

DT 16-JAN-2001 (first entry)
 XX Human APC protein (splice variant 1).
 XX
 KW APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;
 KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;
 KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;
 KW bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;
 KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;
 KW genetic predisposition; drug screening; DP2.5; splice variant.
 XX
 OS Homo sapiens.
 XX
 XX US6114124-A.
 XX
 PD 05-SEP-2000.
 XX
 XX 25-MAY-1995; 95US-00450582.
 XX
 PR 16-JAN-1991; 91GB-00000962.
 PR 16-JAN-1991; 91GB-00000963.
 PR 16-JAN-1991; 91GB-00000974.
 PR 16-JAN-1991; 91GB-00000975.
 PR 08-AUG-1991; 91US-00741940.
 PR 12-AUG-1994; 94US-00289548.
 XX
 XX (ICIL) IMPERIAL CHEM IND PLC.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTAH) UNIV UTAH.
 PA (CANC-) CANCER INST.
 XX
 XX Carlson M, Groden J, Joslyn G, Kinzler K, Markham AP, Anand R;
 PI Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B;
 PI Hedge PJ;
 XX
 DR WPI; 2000-565003/52.
 DR N-PSDB; AAA93449.
 XX
 PT Detecting Adenomatous Polyposis Coli (APC) protein in a sample for
 PT diagnosing cancers, involves contacting the sample with antibodies that
 PT specifically bind to APC protein and detecting the complex formed.
 XX
 PS Claim 1; Fig 3A-C; 125pp; English.
 XX
 CC The invention relates to a novel method for detecting Adenomatous
 CC Polyposis Coli (APC) protein in a sample. The method involves contacting
 CC the sample with antibodies which specifically binds to the 2843 amino
 CC acid form of the human APC protein, or to a mutant APC protein, and
 CC detecting an APC-antibody complex. Mutations in the APC gene play a role
 CC in tumorigenesis, indicating that it is a tumour suppressor gene. It is
 CC located on chromosome 5q21, which corresponds to the FAP (familial
 CC adenomatous polyposis) locus. FAP is an autosomal dominant inherited
 CC disease in which affected individuals develop hundreds to thousands of
 CC adenomatous polyps in the colon and rectum, some of which progress to
 CC malignancy. The FAP locus is often found to be deleted in sporadic (i.e.,
 CC non-familial) adenomas and carcinomas, and chromosome 5q deletions have
 CC also been observed in tumours of the lung, breast, colon, rectum,
 CC bladder, liver, sarcomas, stomach, and prostate, and in leukaemias and
 CC lymphomas. Although the FAP locus contains several other genes such as
 CC FFR, TB1, TB2, and MCC, it is thought that mutations in the APC gene play
 CC a key role in the development of FAP and sporadic tumours. The method is
 CC useful for detecting APC protein and its mutant forms in foetal tissue,
 CC placental tissue, amniotic fluid, blood, serum or a tumour sample. The
 CC method is useful for diagnosing or prognosing neoplastic tissue, for
 CC detecting a genetic predisposition to cancer, for detecting germline and
 CC somatic alteration of wild-type APC genes, and for testing therapeutic
 CC agents for the ability to suppress tumours. The present sequence
 CC represents a 2843 amino acid splice variant of the human APC protein.
 CC This variant is more abundant than the 2742 amino acid variant (AAB23012)
 XX
 XX Sequence 2843 AA;

		Best Local Similarity 99.8%; Pred. No. 0;		Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;	
QY	1	MAAASYDQLLKQVEALKNWENSLNRQLEDNSNHLTKLETEASNKKEVLKQLOGSIEDEAM	60		
DB	1	MAAASYDQLLKQVEALKNWENSLNRQLEDNSNHLTKLETEASNKKEVLKQLOGSIEDEAM	60		
QY	61	ASSQIDLLERLKLNDSSNPFQVKLRQKSRQSRVSGSREGSVSSRGSGCPVPMGSFPR	120		
DB	61	ASSQIDLLERLKLNDSSNPFQVKLRQKSRQSRVSGSREGSVSSRGSGCPVPMGSFPR	120		
QY	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKQWYQAQLQNLTKRIDSLPLTFNPSL	180		
DB	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKQWYQAQLQNLTKRIDSLPLTFNPSL	180		
QY	181	QTDLTRQLEYEARQIRVAMEBQIGTCQDMKEKRAQRRIARIQOIEKDILAIRIOLLOQAT	240		
DB	181	QTDLTRQLEYEARQIRVAMEBQIGTCQDMKEKRAQRRIARIQOIEKDILAIRIOLLOQAT	240		
QY	241	EAERSSQNKHETGSHDAERQEGQVGEINWATSGGQSTTRMDHETASVLSSTHSA	300		
DB	241	EAERSSQNKHETGSHDAERQEGQVGEINWATSGGQSTTRMDHETASVLSSTHSA	300		
QY	301	PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAWSSQDSCINRQSGCPLLIQLL	360		
DB	301	PRRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAWSSQDSCINRQSGCPLLIQLL	360		
QY	361	HGNDKDSVILGNSRGSKEARASAAALHNIHSQDDPKRGREIRVHLLEQIRAYCETC	420		
DB	361	HGNDKDSVILGNSRGSKEARASAAALHNIHSQDDPKRGREIRVHLLEQIRAYCETC	420		
QY	421	WEMQEAHEPGMDQKPNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ	480		
DB	421	WEMQEAHEPGMDQKPNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ	480		
QY	481	VDCMYGLTNDHYSITLRRYAGHALTNLTGVDVANKATLCSCMGKGCRRALVAQIKSSEDL	540		
DB	481	VDCMYGLTNDHYSITLRRYAGHALTNLTGVDVANKATLCSCMGKGCRRALVAQIKSSEDL	540		
QY	541	QOVIASVLRNLSRADVNSKKTJREVGSVKALMECALEVKKESTLKSLSALWLSAHC	600		
DB	541	QOVIASVLRNLSRADVNSKKTJREVGSVKALMECALEVKKESTLKSLSALWLSAHC	600		
QY	601	ENKADI CAVDGAALFVGLTITRSQNTNLAIIESGGILNVSLSIATNEDHRIQIRENN	660		
DB	601	ENKADI CAVDGAALFVGLTITRSQNTNLAIIESGGILNVSLSIATNEDHRIQIRENN	660		
QY	661	CLQTLLOHLKSHSLITVSNACGLTNLSARNPKQDALMDGAVSMKLNLIHSHKXIAM	720		
DB	661	CLQTLLOHLKSHSLITVSNACGLTNLSARNPKQDALMDGAVSMKLNLIHSHKXIAM	720		
QY	721	GSAALRNLMANRPKYKDANINSPGSSLPSSLHVRKQKALEAEALDAQLHSETFDNIDLS	780		
DB	721	GSAALRNLMANRPKYKDANINSPGSSLPSSLHVRKQKALEAEALDAQLHSETFDNIDLS	780		
QY	781	PKASHRSQKHSLGSDYVDFNDRHDDNRSDFNTGNMTVLSPLYNTTVLPSSSSSRGS	840		
DB	781	PKASHRSQKHSLGSDYVDFNDRHDDNRSDFNTGNMTVLSPLYNTTVLPSSSSSRGS	840		
QY	841	LDSSRSKDSLERERIGLGNYPATENPGTSKRGLOISTTTAAQIAKIAMEEVAIHTS	900		
DB	841	LDSSRSKDSLERERIGLGNYPATENPGTSKRGLOISTTTAAQIAKIAMEEVAIHTS	900		
QY	901	QEDRSSSTTELCVTDERNALRRSSAAHSHNTYNTKSENNRCTSCMPYAKLEYKRS	960		
DB	901	QEDRSSSTTELCVTDERNALRRSSAAHSHNTYNTKSENNRCTSCMPYAKLEYKRS	960		
QY	961	NDLSNVSNDGCKRQMKPSIESYSEDDSKFCYGVQPADLAHKIHSANHMDNDGE	1020		
DB	961	NDLSNVSNDGCKRQMKPSIESYSEDDSKFCYGVQPADLAHKIHSANHMDNDGE	1020		
QY	1021	LDTPINTSLKYSBOLNSGRQSPQNERWAPKHIIEDEIKQSEQRNOSTYPTVITE	1080		

Db 1021 LDTPIYSLKYSDBQLNSGRQSPQNERWAPKHHIEDEIKQSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFOHFQOQCVSPYRGRGANGSETNRVGNHGINQVNSQSLQCEDDYEDDKP 1140
Db 1081 STDDKHLKFOHFQOQCVSPYRGRGANGSETNRVGNHGINQVNSQSLQCEDDYEDDKP 1140
Qy 1141 TNYSERVSEBEOHEERPTNYSIKYNEBKXHVDPIDYSLKYATDIPSSQKQSFSEFSK 1200
Db 1141 TNYSERVSEBEOHEERPTNYSIKYNEBKXHVDPIDYSLKYATDIPSSQKQSFSEFSK 1200
Qy 1201 SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Db 1201 SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSLSAASDETCNQTTQEAADSANTLQIAETKIGKIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFSRCSLSLSAASDETCNQTTQEAADSANTLQIAETKIGKIGTRSAEDPV 1320
Qy 1321 SEVPAVSHQPTKSRRLQSSLSSESARHKAVERPSSGAKSPSKCACTPKSPPEHYVOET 1380
Db 1321 SEVPAVSHQPTKSRRLQSSLSSESARHKAVERPSSGAKSPSKCACTPKSPPEHYVOET 1380
Qy 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1440
Db 1381 PLMFSRCTSVSSLSDFSRSTASSVQSEPCSGMVSGIISPSDLDPSPGQTMPPSRSKTTP 1440
Qy 1441 PPPQTAQTKREVPKNAKPTAEKRESGPKQAANVAQVQVLPDADTLLHFAESTPDGF 1500
Db 1441 PPPQTAQTKREVPKNAKPTAEKRESGPKQAANVAQVQVLPDADTLLHFAESTPDGF 1500
Qy 1501 SCSSLSALSALDEFFIIOKVELRIMPVQENDNGNETESEPQKESNENQKEAEKTIIDSE 1560
Db 1501 SCSSLSALSALDEFFIIOKVELRIMPVQENDNGNETESEPQKESNENQKEAEKTIIDSE 1560
Qy 1561 KOLLDDSDDDIEILEECIISAMPTKSRKAKPAQATASKLPPVARKPQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDIEILEECIISAMPTKSRKAKPAQATASKLPPVARKPQLPVYKLLPS 1620
Qy 1621 QNRLOPKHVSFTPGDDMPRYVCEGTPIFSTATSLSDLTIESPPNELAAGEGVRGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPRYVCEGTPIFSTATSLSDLTIESPPNELAAGEGVRGGAQ 1680
Qy 1681 SGEFEKRDITPTEGRSTDEAOGGKTSSTVTELDNDKAEAGDIIAECINAMPKGSKHP 1740
Db 1681 SGEFEKRDITPTEGRSTDEAOGGKTSSTVTELDNDKAEAGDIIAECINAMPKGSKHP 1740
Qy 1741 FRVKIMDOVOOAGASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
Db 1741 FRVKIMDOVOOAGASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNLN 1800
Qy 1801 AERFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSGFAPDSPHHYTPTEGTPYCFSRND 1860
Db 1801 AERFSDNKSQKQNLKNNKDFNDKLPNNEDVRGSGFAPDSPHHYTPTEGTPYCFSRND 1860
Qy 1861 SLSSLDFFDDDDVLSREKAEILKAKENKESAKVTSHTELTSNQQSANKTQATAKQPINR 1920
Db 1861 SLSSLDFFDDDDVLSREKAEILKAKENKESAKVTSHTELTSNQQSANKTQATAKQPINR 1920
Qy 1921 GQPKPILQKSTFPQSSKDIIDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Db 1921 GQPKPILQKSTFPQSSKDIIDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN 1980
Qy 1981 NKENPIKETPPDSQGPSPQAGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
Db 1981 NKENPIKETPPDSQGPSPQAGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDEDLLQ 2040
Qy 2041 ECISAMPKPKKPSRLKGDNEKHSRPNMGGLIGEDLTLDLKDIOKPDSEHGLSPDSNFD 2100
Db 2041 ECISAMPKPKKPSRLKGDNEKHSRPNMGGLIGEDLTLDLKDIOKPDSEHGLSPDSNFD 2100
Qy 2101 WKATQEGANSIVSSLHQAAAAACLSRQASSDSLSLKSGISLGSFPHLTPDQEEKPFT 2160
Db 2101 WKATQEGANSIVSSLHQAAAAACLSRQASSDSLSLKSGISLGSFPHLTPDQEEKPFT 2160

Qy 2161 SNKGPRILKPGKSKTLETYKIESESKGKGGKVKYKSLITGKVRNSSEISQMKQPLOAN 2220
Db 2161 SNKGPRILKPGKSKTLETYKIESESKGKGGKVKYKSLITGKVRNSSEISQMKQPLOAN 2220
Qy 2221 MPISIRGRPTMIHPCGVNRSSSTSPVSKGPKPLKTPASKSPSEGOTATTSPGAKPSVKS 2280
Db 2221 MPISIRGRPTMIHPCGVNRSSSTSPVSKGPKPLKTPASKSPSEGOTATTSPGAKPSVKS 2280
Qy 2281 ELSPVARQTSQIIGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNGISPN 2340
Db 2281 ELSPVARQTSQIIGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNGISPN 2340
Qy 2341 KLSQLPRTSSPSTASTKSSGCKMYTSPGROMSOONLTQOTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSSPSTASTKSSGCKMYTSPGROMSOONLTQOTGLSKNASSIPRSESASKG 2400
Qy 2401 LNMNNGCANKKVELSRMSTKSSGSESDRSERPVLRQSTFIKEAPSPTLRRLKEESA 2460
Db 2401 LNMNNGCANKKVELSRMSTKSSGSESDRSERPVLRQSTFIKEAPSPTLRRLKEESA 2460
Qy 2461 SFESLSPSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWKLPNLSPPTIEYNDG 2520
Db 2461 SFESLSPSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWKLPNLSPPTIEYNDG 2520
Qy 2521 RPAKRHDIAARSHSESPSRPLINRSGTWKREHSHKSSSLPRVSTWRTGSSSSILGASSES 2580
Db 2521 RPAKRHDIAARSHSESPSRPLINRSGTWKREHSHKSSSLPRVSTWRTGSSSSILGASSES 2580
Qy 2581 SPKAKSEDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSTNTSTQTVSSGANGAES 2640
Db 2581 SPKAKSEDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSTNTSTQTVSSGANGAES 2640
Qy 2641 KTLIYQMAFAVSKTEDVWVRIEDCPINNPRSGRSTPTGNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQMAFAVSKTEDVWVRIEDCPINNPRSGRSTPTGNTPPVIDSVSEKANPNIKDSKN 2700
Qy 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVADAPQKQTEIKPGQNNPVPVSETNESPVERT 2760
Db 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVADAPQKQTEIKPGQNNPVPVSETNESPVERT 2760
Qy 2761 PFSSSSSSKHSSPSGTVAAARVTPFNYNPSPRKSSADTSARPSQIPTVNNNTKKRDSKT 2820
Db 2761 PFSSSSSSKHSSPSGTVAAARVTPFNYNPSPRKSSADTSARPSQIPTVNNNTKKRDSKT 2820
Qy 2821 DSTEESGTSQPKRHSGLVTSV 2843
Db 2821 DSTEESGTSQPKRHSGLVTSV 2843
RESULT 9
ABG71105
ID ABG71105 standard; protein; 2843 AA.
XX
AC ABG71105;
XX
DT 08-JAN-2003 (first entry)
XX Human adenomatous polyposis coli (APC) protein.
DE Human: adenomatous polyposis coli; APC; transgenic animal; colon cancer;
KW developmental abnormality; anticancer drug; beta-catenin;
KW Wnt/Wg signalling pathway.
XX Homo sapiens.
XX EP1243646-A2.
XX
PD 25-SEP-2002.
XX
PF 07-MAR-2002; 2002EP-00290571.
XX
PR 19-MAR-2001; 2001US-0276483P.

XX PA (COUL) COUNCIL SCI & IND RES.
XX Bhandari P, Shashidhara IS;
XX WI; 2002-742695/81.
XX New transgenic Drosophila containing the human colon cancer gene
PT Adenomatous Polyposis Coli is useful as an assay model to screen for new
PT drugs, particularly against colon cancer.
XX PS Claim 1; Page 19-20; 52pp; English.
XX CC The invention describes a transgenic Drosophila whose genome comprises
CC the full-length human colon cancer gene Adenomatous Polyposis Coli (APC)
CC allowing regulated mis-expression of the APC gene resulting in
CC developmental abnormalities. The transgenic flies are used to screen and
CC validate efficacy of anticancer drugs, to identify new target proteins
CC interacting with beta-catenin, genes which interact with human APC, to
CC study the biochemical function of human APC and to identify additional
CC components of the Drosophila Wnt/Mg signalling pathway. In particular the
CC flies are used to screen potential drugs against colon cancer. This is
CC the amino acid sequence of the human adenomatous polyposis coli (APC)
CC protein DNA encoding which is incorporated in the transgenic flies
XX Sequence 2843 AA;
Query Match 99.7%; Score 14533; DB 5; Length 2843;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MAAASYDQLLKQVEALKVENSRLRQLEDNSNHLTKLEASNMKEVLKQSGSIEDEAM 60
Db 1 MAAASYDQLLKQVEALKVENSRLRQLEDNSNHLTKLEASNMKEVLKQSGSIEDEAM 60
Qy 61 ASSGQDILLERLKLNDSSNFFGVKLRKMSLRYSVSGSREGSVSRGSCSVPMGSPFR 120
Db 61 ASSGQDILLERLKLNDSSNFFGVKLRKMSLRYSVSGSREGSVSRGSCSVPMGSPFR 120
Qy 121 RGVNFSRGSTGYLEELKERSLLADLQKEEKDQWYQAQNLTKRIDSLPTENFSL 180
Db 121 RGVNFSRGSTGYLEELKERSLLADLQKEEKDQWYQAQNLTKRIDSLPTENFSL 180
Qy 181 QTLTRQLRYEARQIRVAMEEQLGTCQDMKEKAQRRIARIQIQEKDILRIQLQSQAT 240
Db 181 QDTMTRQLRYEARQIRVAMEEQLGTCQDMKEKAQRRIARIQIQEKDILRIQLQSQAT 240
Qy 241 EAERSSQKHETGSHDAERQNGOGVEINMATSNGOGSTTRMDHETASVLSSTSTSA 300
Db 241 EAERSSQKHETGSHDAERQNGOGVEINMATSNGOGSTTRMDHETASVLSSTSTSA 300
Qy 301 PRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMROSGLPLLIQLL 360
Db 301 PRLTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQDSCISMROSGLPLLIQLL 360
Qy 361 HGNDKSVLGNRSGKEARASAAHLNIIHSQDDKGRREIRVLHLEQIRAVCEFC 420
Db 361 HGNDKSVLGNRSGKEARASAAHLNIIHSQDDKGRREIRVLHLEQIRAVCEFC 420
Qy 421 WENQEAHEPGMDQKNFMPAPVEHQICPAVCVLMKLSFDEEHRHAMNGLGQIAELLQ 480
Db 421 WENQEAHEPGMDQKNFMPAPVEHQICPAVCVLMKLSFDEEHRHAMNGLGQIAELLQ 480
Qy 481 VDCVMYGLTNDHYSITLRRVAGVALTNLTFGVANKATLCMKGCGRVALVAQKSESDL 540
Db 481 VDCVMYGLTNDHYSITLRRVAGVALTNLTFGVANKATLCMKGCGRVALVAQKSESDL 540
Qy 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSLSALWNSAHC 600
Db 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSLSALWNSAHC 600
Qy 601 ENKADICAVDGCALAFVLGTLTYRSQNTNLAIIBSGGGILRNVSLLIATNEDHQILRENN 660
Db 601 ENKADICAVDGCALAFVLGTLTYRSQNTNLAIIBSGGGILRNVSLLIATNEDHQILRENN 660

Db 601 ENKADICAVDGCALAFVLGTLTYRSQNTNLAIIBSGGGILRNVSLLIATNEDHQILRENN 660
Qy 661 CLQTLQHLKSHSLTIIVSNACGTLWNLNARPKDQEAALWDMGAVSMKLNLIHSHKXIAM 720
Db 661 CLQTLQHLKSHSLTIIVSNACGTLWNLNARPKDQEAALWDMGAVSMKLNLIHSHKXIAM 720
Qy 721 GSAALRNLMANRPAPKADANIMSPGSSLSLHVRCOKALEAELDAQHLSSETDNDLS 780
Db 721 GSAALRNLMANRPAPKADANIMSPGSSLSLHVRCOKALEAELDAQHLSSETDNDLS 780
Qy 781 FKASHRSKQRHKQSLGYDYVDTNRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSRGS 840
Db 781 FKASHRSKQRHKQSLGYDYVDTNRHDDNRSDFNTGNMTVLSPYLNTTVLPSSSSRGS 840
Qy 841 LDSRSRSDKSLERERGIGLGNYPATENGTSSKRGLOISTTAAQIAKWEVSIAHTS 900
Db 841 LDSRSRSDKSLERERGIGLGNYPATENGTSSKRGLOISTTAAQIAKWEVSIAHTS 900
Qy 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSTHTYNTFTKSENSNRITCSPYAKLEYKRSS 960
Db 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSTHTYNTFTKSENSNRITCSPYAKLEYKRSS 960
Qy 961 NDSLNSVSSNDGYKRGQMKPSTESYSEDDSKFCYGYQYPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLNSVSSNDGYKRGQMKPSTESYSEDDSKFCYGYQYPADLAHKIHSANHMDNDGE 1020
Qy 1021 LDTPIYSLKYSDQLNSGRQSPSONERWARPKHIIIDEIKQSEQRQSRNOSTTYPVTE 1080
Db 1021 LDTPIYSLKYSDQLNSGRQSPSONERWARPKHIIIDEIKQSEQRQSRNOSTTYPVTE 1080
Qy 1081 STDDKHLKFPQHFQCCVSPYRSGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140
Db 1081 STDDKHLKFPQHFQCCVSPYRSGANGSETNRVGNHGINQVNSOSLQCEDDYEDDKP 1140
Qy 1141 TNYSERISEEHEEERPTNYSIKYNEERHVDQPIDYSLKYATDIPSSQKOSFSEFSKS 1200
Db 1141 TNYSERISEEHEEERPTNYSIKYNEERHVDQPIDYSLKYATDIPSSQKOSFSEFSKS 1200
Qy 1201 SSGQSKTEHMSSESTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
Db 1201 SSGQSKTEHMSSESTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFSCSSLSLSAEDDEIGCNOTTQEAADSANTLOIAETKIGKIGRSADPV 1320
Db 1261 TYCVEDTPICFSCSSLSLSAEDDEIGCNOTTQEAADSANTLOIAETKIGKIGRSADPV 1320
Qy 1321 SEVPAVQHPRTSSRLQSSLSASARHKAVERFPSSGAKSPSKSGAQTPKSPPEHYVQET 1380
Db 1321 SEVPAVQHPRTSSRLQSSLSASARHKAVERFPSSGAKSPSKSGAQTPKSPPEHYVQET 1380
Qy 1381 PLMSRCTSVSSLDSPESRSIASVQSEPCSGVSGIISPSDLPDPSGQTPMPSRSKTPP 1440
Db 1381 PLMSRCTSVSSLDSPESRSIASVQSEPCSGVSGIISPSDLPDPSGQTPMPSRSKTPP 1440
Qy 1441 PPTQATQKREVPKNAKPTAEKRESGPKQAANAAVQVQVLPDADTLHLHFATESPDGF 1500
Db 1441 PPTQATQKREVPKNAKPTAEKRESGPKQAANAAVQVQVLPDADTLHLHFATESPDGF 1500
Qy 1501 SCSSSLSALSDEPFIQKDVLEIRIMPPVQNDNGNETESEPQKESNENQKEAKIIDSE 1560
Db 1501 SCSSSLSALSDEPFIQKDVLEIRIMPPVQNDNGNETESEPQKESNENQKEAKIIDSE 1560
Qy 1561 KDLDDDDDDDDIEILEBICIISAMPTKSSRKPKPAOTASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KDLDDDDDDDDIEILEBICIISAMPTKSSRKPKPAOTASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOPKHVSFTPGDDMPRYCVVEGTPINFSTATSLDITIESSPNVELAAGVGVGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPRYCVVEGTPINFSTATSLDITIESSPNVELAAGVGVGGAQ 1680
Qy 1681 SGFEKEDTIPTEGRSTDEAGGKTSSTVTPPELDDNKABEGDILAEICINSAMPKGSHPK 1740
Db 1681 SGFEKEDTIPTEGRSTDEAGGKTSSTVTPPELDDNKABEGDILAEICINSAMPKGSHPK 1740

QY 1741 FRVKIMDQVOQASASSAPKNQDGGKKKPTSPVKPIQNTYTRVRKNADSKNNLN 1800
 Db 1741 FRVKIMDQVOQASASSAPKNQDGGKKKPTSPVKPIQNTYTRVRKNADSKNNLN 1800
 QY 1801 AERFVNKSKONLKNKNDKLPNNEDVRGSGFADSPSHYPTIEGTVCFSRND 1860
 Db 1801 AERFVNKSKONLKNKNDKLPNNEDVRGSGFADSPSHYPTIEGTVCFSRND 1860
 QY 1861 SLSLDFDDDDVDLSREKAEKAKENKSEAKVTSHTLTSNQOSANKTQAIKQPINR 1920
 Db 1861 SLSLDFDDDDVDLSREKAEKAKENKSEAKVTSHTLTSNQOSANKTQAIKQPINR 1920
 QY 1921 GQPKILQKSTFPQSSKDIPDRGAATDEKLOFAIENTPVCFSHNSLSLSIDIOENN 1980
 Db 1921 GQPKILQKSTFPQSSKDIPDRGAATDEKLOFAIENTPVCFSHNSLSLSIDIOENN 1980
 QY 1981 NKNEPIKETEPDPSQGEPSKQAGVAPKSFHVEDTPVCFSRNSSLSSLSIDSEDDLQ 2040
 Db 1981 NKNEPIKETEPDPSQGEPSKQAGVAPKSFHVEDTPVCFSRNSSLSSLSIDSEDDLQ 2040
 QY 2041 ECISAMPKKKPKRLKDNKXKSPRNMGGILGEDLTLDKDIORPSEHGLSPDSENF 2100
 Db 2041 ECISAMPKKKPKRLKDNKXKSPRNMGGILGEDLTLDKDIORPSEHGLSPDSENF 2100
 QY 2101 WKAIQEGANSVSSLHQAACLSROASSDSLSILKSGISLGSPPHLPDQEEKPFT 2160
 Db 2101 WKAIQEGANSVSSLHQAACLSROASSDSLSILKSGISLGSPPHLPDQEEKPFT 2160
 QY 2161 SNKGPRILKPEKSTLETTKIESBSKGIKGGKVKYKSLITGVKRSNSEISQMKQPIQAN 2220
 Db 2161 SNKGPRILKPEKSTLETTKIESBSKGIKGGKVKYKSLITGVKRSNSEISQMKQPIQAN 2220
 QY 2221 MPSISRGRTMIHIPVRNNSSTSPVSKGPKPLTPASKSPSEGTATTPRGAKPVSKS 2280
 Db 2221 MPSISRGRTMIHIPVRNNSSTSPVSKGPKPLTPASKSPSEGTATTPRGAKPVSKS 2280
 QY 2281 ELSPVARTSITGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPCRNSISPRNGISPPN 2340
 Db 2281 ELSPVARTSITGSSKAPSRSGRSDTSPRPAQOPLSRPIQSPCRNSISPRNGISPPN 2340
 QY 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMSQOQLTKOTGLSKNASSIPRESASKG 2400
 Db 2341 KLSQLPRTSSPTASTKSGSGKMSYTPGQMSQOQLTKOTGLSKNASSIPRESASKG 2400
 QY 2401 LNMNMNGANKKVLNMSSTKSSGSDSDRSEPVLRQSTFFKEAPSPILRKLBEA 2460
 Db 2401 LNMNMNGANKKVLNMSSTKSSGSDSDRSEPVLRQSTFFKEAPSPILRKLBEA 2460
 QY 2461 SPESLSPSRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPIEYNDG 2520
 Db 2461 SPESLSPSRPASPTRSQAQTPVLSPLPDMSLSTHSSVQAGGWRKLPNLSPIEYNDG 2520
 QY 2521 RPAKRHDIAHSBSPRLPINRSGTWKREHSHSSLPVSTWTRTSSSSSILSASSES 2580
 Db 2521 RPAKRHDIAHSBSPRLPINRSGTWKREHSHSSLPVSTWTRTSSSSSILSASSES 2580
 QY 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNTSOTVSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFPSTNTSOTVSSGATNGAES 2640
 QY 2641 KTLIYQMAVASKTEEDVWVRIEDCPINPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 Db 2641 KTLIYQMAVASKTEEDVWVRIEDCPINPRSGRSPGTNTPPVIDSVSEKANPNIKSDKN 2700
 QY 2701 QAKQNVGNSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNVPVSETNESPIVERT 2760
 Db 2701 QAKQNVGNSVPMRTVGLNRLTSTFIQVDAPQKGTETKPGQNNVPVSETNESPIVERT 2760
 QY 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPFRKSSADTSARSQIPTPVNNNTKRDST 2820
 Db 2761 PFSSSSSSKHSPSGTVAARVTPFNPNPFRKSSADTSARSQIPTPVNNNTKRDST 2820

QY 2821 DSTESSTQSPKRHSGSYLTVTSV 2843
 Db 2821 DSTESSTQSPKRHSGSYLTVTSV 2843
 RESULT 10
 AAW76821
 ID AAW76821 standard; protein; 2973 AA.
 XX AAW76821;
 XX AC
 XX XX
 DT 25-JAN-1999 (first entry)
 XX
 DE Human APC protein.
 XX
 XX Tcf; beta-catenin; human; drug; familial adenomatous polyposis; FAP;
 KW cancer; adenomatous polyposis coli; APC; neoplastic.
 XX
 OS Homo sapiens.
 XX
 PN WC9841631-A2.
 XX
 PD 24-SEP-1998.
 XX
 PF 20-MAR-1998; 98WO-US005506.
 XX
 PR 20-MAR-1997; 97US-00821355.
 XX
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UYUT-) RIJKSUNIV UTRECHT.
 XX
 PI Barker N, Clevers H, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
 PI Vogelstein B;
 PI
 DR WPI; 1998-531569/45.
 XX
 XX
 PT Intron-free DNA encoding Tcf-4 protein - useful for, e.g. identifying
 PT drugs for treating FAP patients, or patients with increased risk of
 XX developing cancer.
 PS Disclosure; Page 32-37; 58pp; English.
 XX
 CC This sequence represents a human APC protein which is used in a method of
 CC identifying candidate drugs for use in familial adenomatous polyposis
 CC (FAP) patients, or patients with increased risk of developing cancer. The
 CC protein can also be used to determine the presence or absence in a cell
 CC of wild type adenomatous polyposis coli (APC) gene or a downstream
 CC protein in the APC transcription regulatory pathway. This method involves
 CC introducing a Tcf-responsive reporter gene into the cell, and measuring
 CC transcription of the reporter gene where a cell which supports active
 CC downstream protein of the APC transcription regulatory pathway. The
 CC protein can also be used in a method of diagnosing cancer in a sample
 CC suspected of being neoplastic and for treating a patient with colorectal
 CC cancer or other cancer associated with FAP, comprising administering to
 CC the patient a polypeptide comprising a portion of the APC sequence,
 CC comprising the beta-catenin binding site
 XX
 SQ Sequence 2973 AA;
 Query Match 99.7%; Score 14533; DB 2; Length 2973;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAAASYDQLLKQVEALKMENSLRQELDNDNSNLTLEASNNKVKLQKQSIIDEAM 60
 Db 1 MAAASYDQLLKQVEALKMENSLRQELDNDNSNLTLEASNNKVKLQKQSIIDEAM 60
 QY 61 ASSQIDILLERLKEINLDSNFPVGVKLRKWSLRYSRSGSVSSRSGECPVPMGSPFR 120
 Db 61 ASSQIDILLERLKEINLDSNFPVGVKLRKWSLRYSRSGSVSSRSGECPVPMGSPFR 120
 QY 121 RGVNNGRSTGYLLEEKERSLLALLDKKEEKEKWYIAQLQNLTKRISLPLTENFSL 180

Db 121 RGVNGRESYGLLELEKERSLLADLDEEKEKWYAYQLOQLTKRIDSLPLTENFSL 180
Qy 181 QTLTRRQLEYEARQIRVAMEEQLGTQDMKKAQRRIARIQIIEKDIIRIROLLSQAT 240
Db 181 QTDWTRRQLEYEARQIRVAMEEQLGTQDMKKAQRRIARIQIIEKDIIRIROLLSQAT 240
Qy 241 EABRSSQKHETGSHDAERONEGGVGEINWATSGNGQGGTTRMDHETASVLSSTHSA 300
Db 241 EABRSSQKHETGSHDAERONEGGVGEINWATSGNGQGGTTRMDHETASVLSSTHSA 300
Qy 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDDMSRTLLAMSSQDSCIENRSGGCLPLIOLL 360
Db 301 PRRLTSHLGTKEVMYSLLSMLGTHDKDDMSRTLLAMSSQDSCIENRSGGCLPLIOLL 360
Qy 361 HGNDKDSVLLGNRSGSKEAPARASAAALHNTIHSQDDKGRREIRVJLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNRSGSKEAPARASAAALHNTIHSQDDKGRREIRVJLHLEQIRAYCETC 420
Qy 421 WEMQEAHEPGWODKNPMPAPVHOICPAVCVLMKLSFDEEHRHAMVELGLOAIAELLO 480
Db 421 WEMQEAHEPGWODKNPMPAPVHOICPAVCVLMKLSFDEEHRHAMVELGLOAIAELLO 480
Qy 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTGVDVANKATLCSMKGCMRALVAQLKSESED 540
Db 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTGVDVANKATLCSMKGCMRALVAQLKSESED 540
Qy 541 QOVTASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVLISALWNLISAKCT 600
Db 541 QOVTASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVLISALWNLISAKCT 600
Qy 601 ENKADICAVDGAFLAVGLTITYSQNTLAIIBSGGILRNYSLLIATNEDHQILRENN 660
Db 601 ENKADICAVDGAFLAVGLTITYSQNTLAIIBSGGILRNYSLLIATNEDHQILRENN 660
Qy 661 CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALMDMGAVSMKMLIHSKHKQIAM 720
Db 661 CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALMDMGAVSMKMLIHSKHKQIAM 720
Qy 721 GSAAALRNLMANRPKAKDANIMSGSILPSLHVKKOKALEAELDAHLSETFDINDLS 780
Db 721 GSAAALRNLMANRPKAKDANIMSGSILPSLHVKKOKALEAELDAHLSETFDINDLS 780
Qy 781 PKASHRSKORHKQSLYGVYVFTNRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSRGS 840
Db 781 PKASHRSKORHKQSLYGVYVFTNRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSRGS 840
Qy 841 LDSRSRSEKORSLERRGIGLVNHPATENPGTSSKGLQISTTAQIAKVMEEVSAIHTS 900
Db 841 LDSRSRSEKORSLERRGIGLVNHPATENPGTSSKGLQISTTAQIAKVMEEVSAIHTS 900
Qy 901 QEDRSSGSTTELHCVTDERNALRRSSAAHTSNFTKSENSTSCMPYAKLEYKSS 960
Db 901 QEDRSSGSTTELHCVTDERNALRRSSAAHTSNFTKSENSTSCMPYAKLEYKSS 960
Qy 961 NDSLSNSVNDGCVKQGMKPSIESSEDESKFSYGOYPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLSNSVNDGCVKQGMKPSIESSEDESKFSYGOYPADLAHKIHSANHMDNDGE 1020
Qy 1021 LDTPIINYSKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQNRQSTTYPVYTE 1080
Db 1021 LDTPIINYSKYDEQLNSGRQSPQNERWARPKHIEDEIKQSEQRQNRQSTTYPVYTE 1080
Qy 1081 STDDKHLKOPHFQOECVPSYRSGANGSETRVGSNHNQNVQSILCOEDDYEDDKP 1140
Db 1081 STDDKHLKOPHFQOECVPSYRSGANGSETRVGSNHNQNVQSILCOEDDYEDDKP 1140
Qy 1141 TNYSERYSSEEEHEEERTNYSIKNEEKRVADPIDYSLKYATDIPSSOKQSPFSKS 1200
Db 1141 TNYSERYSSEEEHEEERTNYSIKNEEKRVADPIDYSLKYATDIPSSOKQSPFSKS 1200
Qy 1201 SSGQSKTEHMSSENSTTPSSNAKRONQLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260

Db 1201 SSGQSKTEHMSSENSTTPSSNAKRONQLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTTPICFRSCSSLSLSAARDEICNCTTOEADSANTLOIAEIKGKTGTSADDPV 1320
Db 1261 TYCVEDTTPICFRSCSSLSLSAARDEICNCTTOEADSANTLOIAEIKGKTGTSADDPV 1320
Qy 1321 SEVPAVSQHPRTKSRLOGSSLSSESARHKAVERFPSSGAKSPSKGAOTPKSPPHYVQET 1380
Db 1321 SEVPAVSQHPRTKSRLOGSSLSSESARHKAVERFPSSGAKSPSKGAOTPKSPPHYVQET 1380
Qy 1381 PLMFSECTSVSSLOSFSRSIASSVQSEPCSGMVGIIISPSDLPSDQOTMPPRSKTPP 1440
Db 1381 PLMFSECTSVSSLOSFSRSIASSVQSEPCSGMVGIIISPSDLPSDQOTMPPRSKTPP 1440
Qy 1441 PPPOTAQTKREVPKNKAFTAEKRESGPKQAANAAVQVQLPADTLLHFAFTESTPDGF 1500
Db 1441 PPPOTAQTKREVPKNKAFTAEKRESGPKQAANAAVQVQLPADTLLHFAFTESTPDGF 1500
Qy 1501 SCSSLSLSALDEPFIQKDVLELRIMPPVQENDNGNETEPOPKESNENQKEAEKTIIDSE 1560
Db 1501 SCSSLSLSALDEPFIQKDVLELRIMPPVQENDNGNETEPOPKESNENQKEAEKTIIDSE 1560
Qy 1561 KDLDDSDDDDDIEILEECIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KDLDDSDDDDDIEILEECIISAMPTKSRKPKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLSDLTIESPPNELAAGEVGRGAQ 1680
Db 1621 QNRLOQKHVSFTPGDDMPRVYCVGEGTPIINFSTATISLSDLTIESPPNELAAGEVGRGAQ 1680
Qy 1681 SGEFEKDDTTPTEGRSTDEAGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKGSHP 1740
Db 1681 SGEFEKDDTTPTEGRSTDEAGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKGSHP 1740
Qy 1741 FRVKIMQVQOQASASSAPNKNQLDGKKKPTSPVKPIPONTTEYRTRVRKNADSKNLN 1800
Db 1741 FRVKIMQVQOQASASSAPNKNQLDGKKKPTSPVKPIPONTTEYRTRVRKNADSKNLN 1800
Qy 1801 AERFVSNDKSKONKNNSKDNFNDKLPNNEDVRGSPAFDSPHHTPIEGTGYCFSRND 1860
Db 1801 AERFVSNDKSKONKNNSKDNFNDKLPNNEDVRGSPAFDSPHHTPIEGTGYCFSRND 1860
Qy 1861 SLSLSDDDDDVLSREKAEELKAKENKESEAKVTSHTELTSNQOSANKTQAIKAPINR 1920
Db 1861 SLSLSDDDDDVLSREKAEELKAKENKESEAKVTSHTELTSNQOSANKTQAIKAPINR 1920
Qy 1921 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLONFALIENTPVCFSHNSSLSLSDIDQENN 1980
Db 1921 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLONFALIENTPVCFSHNSSLSLSDIDQENN 1980
Qy 1981 NKNEPIKETEPDSDGCEPSKPOASGYAPKSFHVEDTPTVCFSRNSSLSLSDIDQENN 2040
Db 1981 NKNEPIKETEPDSDGCEPSKPOASGYAPKSFHVEDTPTVCFSRNSSLSLSDIDQENN 2040
Qy 2041 ECISAMPKKKPSRLKGDNEKHSPRNMGILGEDTLDLKDIQRPDSEHGLSPDSENF 2100
Db 2041 ECISAMPKKKPSRLKGDNEKHSPRNMGILGEDTLDLKDIQRPDSEHGLSPDSENF 2100
Qy 2101 WKAIQEGANSIVSLSHQAAAACLSRQASDSDSILSLKSGISLGSFPHLTPDQEKPF 2160
Db 2101 WKAIQEGANSIVSLSHQAAAACLSRQASDSDSILSLKSGISLGSFPHLTPDQEKPF 2160
Qy 2161 SNKGPRIKPEKSTLETKKIESEKIGKGVKYSKITGKVRNSSETSGQMKQPLQAN 2220
Db 2161 SNKGPRIKPEKSTLETKKIESEKIGKGVKYSKITGKVRNSSETSGQMKQPLQAN 2220
Qy 2221 MPSISRGRTWIHPGVNRNSSSTSPYKGGPPLKTPASKSPSEGOATTPSKGKPSVKS 2280
Db 2221 MPSISRGRTWIHPGVNRNSSSTSPYKGGPPLKTPASKSPSEGOATTPSKGKPSVKS 2280
Qy 2281 ELSFVARQTSIQGSSKAPSRGSRDSTPSRPAQOPLSRPIQSPGRNISIPGRNGISPPN 2340
Db 2281 ELSFVARQTSIQGSSKAPSRGSRDSTPSRPAQOPLSRPIQSPGRNISIPGRNGISPPN 2340

Qy 2341 KLSOLPRTSPSTASTKSSGKWSYTSPCROMSQOQLTKOTGLSKNASSIPESASKG 2400
 Db 2341 KLSOLPRTSPSTASTKSSGKWSYTSPCROMSQOQLTKOTGLSKNASSIPESASKG 2400
 Qy 2401 LNMWNGGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA 2460
 Db 2401 LNMWNGGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA 2460
 Qy 2461 SFESLSPSPASTRQAQTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEVNDG 2520
 Db 2461 SFESLSPSPASTRQAQTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEVNDG 2520
 Qy 2521 RPAKHDIARSHSPSRPLNRSRGTWKREHSHKSSSLPRVSTWRTTGGSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSPSRPLNRSRGTWKREHSHKSSSLPRVSTWRTTGGSSSSILSASSES 2580
 Qy 2581 SEKAKEDEKHVNSISGTSKQENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Db 2581 SEKAKEDEKHVNSISGTSKQENQVSAKGTWRKIKENEFSPNTSTQTVSSGATNGAES 2640
 Qy 2641 KTLIYQMAPAVSKTQEDVWRIEDCPINNPRSGRPTGNTFPVIDSVSEKANPNIKOSKON 2700
 Db 2641 KTLIYQMAPAVSKTQEDVWRIEDCPINNPRSGRPTGNTFPVIDSVSEKANPNIKOSKON 2700
 Qy 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVDAPDQKTEIKPQNNPVPVSETNESPIVERT 2760
 Db 2701 QAKQNVGNGSVPMRTVGLNRLTSFIQVDAPDQKTEIKPQNNPVPVSETNESPIVERT 2760
 Qy 2761 PFSSSSSKHSPSGTVAARVTPPNVNSPKSSADSTSAEPSOIPTPVNNTKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARVTPPNVNSPKSSADSTSAEPSOIPTPVNNTKRDSKT 2820
 Qy 2821 DSTESSGTQSPKRHSGSYLTVSV 2843
 Db 2821 DSTESSGTQSPKRHSGSYLTVSV 2843
 RESULT 11
 ID AAY72782
 XX AAY72782 standard; protein; 2973 AA.
 AC AAY72782;
 XX AAY72782;
 DT 31-MAY-2001 (first entry)
 XX Transcriptional activation protein #1 related to the invention.
 DE Human; Tcf-4 protein; transcriptional activation factor; beta-catenin;
 KW apoptosis; colorectal cancer; Familial Adenomatous Polyposis; FAP; APC;
 KW adenomatous polyposis coli; recombinant adenovirus; Ad-Mini-Me; therapy;
 KW GFP; green fluorescent protein; GFP/CAPC fusion protein; cytosstatic.
 XX Homo sapiens.
 OS WO200116167-A2.
 PN 08-MAR-2001.
 XX 29-AUG-2000; 2000WO-US023635.
 PF 01-SEP-1999; 99US-00388354.
 PR (UJO) UNIV JOHNS HOPKINS.
 PA Barker N, Clevers JC, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
 PI Vogelstein B, He T;
 XX WPI; 2001-226675/23.
 DR New fusion protein comprising an enzyme covalently linked to a portion of
 PT the adenomatous polyposis coli comprising its beta-catenin binding domain
 PT useful for inducing apoptosis or treating colorectal cancer.

XX Disclosure; Page 74-81; 83pp; English.
 PS The present sequence is a transcriptional activation protein related to
 XX the invention. The invention relates to human Tcf-4 proteins and their
 CC corresponding cDNA molecules which encodes transcriptional activation
 CC factors of human Tcf/Lef family. Human Tcf-4 binds to beta-catenin and
 CC activates transcription in colorectal epithelial cells. Moreover it has
 CC been found that adenomatous polyposis coli (APC) regulates this
 CC transcriptional activation, by binding to beta-catenin. The invention
 CC also provides a recombinant adenovirus, Ad-Mini-Me ie., APC Minus its
 CC amino- and carboxyl- terminal ends which expresses a fusion protein,
 CC green fluorescent protein (GFP)/CAPC containing GFP fused to the central
 CC third of APC which contains its beta-catenin binding domain useful for
 CC inducing apoptosis or treating colorectal cancer. These fusion proteins
 CC are useful for treating cancer, e.g. colorectal cancer, and other cancers
 CC associated with Familial Adenomatous Polyposis (FAP) or patients with
 CC increased risk of developing cancer. Human Tcf-4 cDNA provides an
 CC excellent system for screening agents for their ability to promote
 CC delivery, integration, hybridization, expression, replication or
 CC integration in cells or in an animal. It also provides methods for
 CC diagnosis cancer in a sample suspected of being neoplastic
 XX
 SQ Sequence 2973 AA;
 Query Match 99.7%; Score 14533; DB 4; Length 2973;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Qy 1 MAAASYDQLLKQVEALKMENSRLROELEDNSNHLTKLETEASNKEVLKQLOGSIEDEAM 60
 Db 1 MAAASYDQLLKQVEALKMENSRLROELEDNSNHLTKLETEASNKEVLKQLOGSIEDEAM 60
 Qy 61 ASSGQIDLLERLKLNDSSNPFVKLRSMKLSIRSYGSRGSGSVSRGSCSPVPMGSPFR 120
 Db 61 ASSGQIDLLERLKLNDSSNPFVKLRSMKLSIRSYGSRGSGSVSRGSCSPVPMGSPFR 120
 Qy 121 RGFVNGSRSTGYLELEKERSILLADLDKEKEKDWYAYQNLTKRIDSPLTENFSL 180
 Db 121 RGFVNGSRSTGYLELEKERSILLADLDKEKEKDWYAYQNLTKRIDSPLTENFSL 180
 Qy 181 QDLTRQLLEYEARQIRVAMEBOLGTCQDMKEKRAQRIARIQOIEKDIIRQLLOSQT 240
 Db 181 QDWTTRQLLEYEARQIRVAMEEQLGTCQDMKEKRAQRIARIQOIEKDIIRQLLOSQT 240
 Qy 241 EARSQNHETGSHDAERQNEGGVGEINMATSNGQSGTTRMDHETASVLSSSSTHSA 300
 Db 241 EARSQNHETGSHDAERQNEGGVGEINMATSNGQSGTTRMDHETASVLSSSSTHSA 300
 Qy 301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSSQSCISMRQSGCLPLLIQLL 360
 Db 301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDMSRTLLAMSSSQSCISMRQSGCLPLLIQLL 360
 Qy 361 HGNDKDSVLLGNRSGSKEARASAAHLNTHSQDDPKGRREIRVHLLEQIRAYCETC 420
 Db 361 HGNDKDSVLLGNRSGSKEARASAAHLNTHSQDDPKGRREIRVHLLEQIRAYCETC 420
 Qy 421 WEWQEAHEPGMDQDKNPMAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
 Db 421 WEWQEAHEPGMDQDKNPMAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQAIALLQ 480
 Qy 481 VDCMYGLTNDHYSITLRRYAGMALTNLPGDVANKATLCSMKGCMRALVAQKSEEDL 540
 Db 481 VDCMYGLTNDHYSITLRRYAGMALTNLPGDVANKATLCSMKGCMRALVAQKSEEDL 540
 Qy 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHCT 600
 Db 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHCT 600
 Qy 601 ENKADI CAVDGALAFVLGTITRYSQNTNLTALIESGGGILRNVSLSLATNEDHRLRENN 660
 Db 601 ENKADI CAVDGALAFVLGTITRYSQNTNLTALIESGGGILRNVSLSLATNEDHRLRENN 660

QY 661 CLOTLQHLKSHSLTIYSNACGTLWNLARNPKDQEBALWDMGAVSMLKNLIHSHKHMIAM 720
Db 661 CLOTLQHLKSHSLTIYSNACGTLWNLARNPKDQEBALWDMGAVSMLKNLIHSHKHMIAM 720
QY 721 GSAALANLWANRPKAKYDANIMSPGSLPSLHVHRKQKALEABELDAQHSETFDNDNL 780
Db 721 GSAALANLWANRPKAKYDANIMSPGSLPSLHVHRKQKALEABELDAQHSETFDNDNL 780
QY 781 PKSHRSKORHKSLYGDYVFDNRHDDNRSDFNTGNMTVLSPYLTNTVLPSSSSRGS 840
Db 781 PKSHRSKORHKSLYGDYVFDNRHDDNRSDFNTGNMTVLSPYLTNTVLPSSSSRGS 840
QY 841 LQSSRSKRLERERIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS 900
Db 841 LQSSRSKRLERERIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS 900
QY 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSHNTYNTFTKSENSNRCTSMPIAKLEYKRSS 960
Db 901 QEDRSSGTTTELHCVTDERNALRRSSAAHTSHNTYNTFTKSENSNRCTSMPIAKLEYKRSS 960
QY 961 NDSLSYSSNDGKGRQOMKPSIESYSEDDDESDFCSYQVQPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLSYSSNDGKGRQOMKPSIESYSEDDDESDFCSYQVQPADLAHKIHSANHMDNDGE 1020
QY 1021 LDPINYSKYDEQLNSGRQSPQNERWARPKHIIIDEIKQEQQRQSNQSTTTPVYTE 1080
Db 1021 LDPINYSKYDEQLNSGRQSPQNERWARPKHIIIDEIKQEQQRQSNQSTTTPVYTE 1080
QY 1081 STDDKHLKFQHPHQEQCVSPYRSRGANGSETNRVGSNGINQNVOSQICQEDDDDDKP 1140
Db 1081 STDDKHLKFQHPHQEQCVSPYRSRGANGSETNRVGSNGINQNVOSQICQEDDDDDKP 1140
QY 1141 TNSERYSEBEEHEBERPNTYIKYNEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
Db 1141 TNSERYSEBEEHEBERPNTYIKYNEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
QY 1201 SSGQSKTEHMSSESENTSTPSSNAKRONQLHPSSAQSRSGOPKAACTKVSSINQETIQ 1260
Db 1201 SSGQSKTEHMSSESENTSTPSSNAKRONQLHPSSAQSRSGOPKAACTKVSSINQETIQ 1260
QY 1261 TYCVETPTICFSCSSLSLSAEDIEGNCQTQEADSANTLQIAIEIKIGITRGAEDPV 1320
Db 1261 TYCVETPTICFSCSSLSLSAEDIEGNCQTQEADSANTLQIAIEIKIGITRGAEDPV 1320
QY 1321 SEVPAYSQHPTKSRRLQSSLSSESARKVAFPPSGAKSPKSGAQTPKSPPEHYVQET 1380
Db 1321 SEVPAYSQHPTKSRRLQSSLSSESARKVAFPPSGAKSPKSGAQTPKSPPEHYVQET 1380
QY 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPSPGOTMPPSRSKTTP 1440
Db 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPSPGOTMPPSRSKTTP 1440
QY 1441 PPQTQATKREVPKNAKPAETAKRESQPKAAVNAVQVLPDADTLHFAETSTPDGF 1500
Db 1441 PPQTQATKREVPKNAKPAETAKRESQPKAAVNAVQVLPDADTLHFAETSTPDGF 1500
QY 1501 SCSSLSLSALSLDEPFIQKQVELRIMPVQENDNGNETESEOKESENOEKEAETIDSE 1560
Db 1501 SCSSLSLSALSLDEPFIQKQVELRIMPVQENDNGNETESEOKESENOEKEAETIDSE 1560
QY 1561 KOLLDDSDDDDEILBECIISAMPKSKSRGKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDDEILBECIISAMPKSKSRGKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
QY 1621 QNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLSDLTIESPNELAAAGEVRGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPRVYCVETGPIINFSTATSLSDLTIESPNELAAAGEVRGGAQ 1680
QY 1681 SGEFEKRDITPEGRSTDBAQGKTSSTVITPELDDNKAPEGDILAEICINSAMPKSKHP 1740
Db 1681 SGEFEKRDITPEGRSTDBAQGKTSSTVITPELDDNKAPEGDILAEICINSAMPKSKHP 1740
QY 1741 FRVVKIMDQVOQASASSAPNKNQDGGKKKPTSPVKPIPQNTYRTRVRKNADSKNNLN 1800

Db 1741 FRVVKIMDQVOQASASSAPNKNQDGGKKKPTSPVKPIPQNTYRTRVRKNADSKNNLN 1800
QY 1801 AERVPFNDKOSKQNLKXNSKDFNDKLPNNEDRVRSFAFDSPPHYTPIEGTYPYCFGRND 1860
Db 1801 AERVPFNDKOSKQNLKXNSKDFNDKLPNNEDRVRSFAFDSPPHYTPIEGTYPYCFGRND 1860
QY 1861 SLSLDPDDDDVDLSREKAEIRKAKENKESFAKTSHTELTSNQQSANKTAQAKQPINR 1920
Db 1861 SLSLDPDDDDVDLSREKAEIRKAKENKESFAKTSHTELTSNQQSANKTAQAKQPINR 1920
QY 1921 GQPKPILOKQSTFPQSSKDIPIRGAATDEKLQFAIENTPVCFSHNSLSLSLSDIDENN 1980
Db 1921 GQPKPILOKQSTFPQSSKDIPIRGAATDEKLQFAIENTPVCFSHNSLSLSLSDIDENN 1980
QY 1981 NKENPIKETEPBPPSQEPSPKQASGVAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
Db 1981 NKENPIKETEPBPPSQEPSPKQASGVAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
QY 2041 ECISAMPKPKKPRKRLKGDNEKHSRNNMGILGEBDLTLDLKDITORPDSHGLSPDSENF 2100
Db 2041 ECISAMPKPKKPRKRLKGDNEKHSRNNMGILGEBDLTLDLKDITORPDSHGLSPDSENF 2100
QY 2101 WKAIQEGANSIVSLHQAACAAACLSRQASDSDSIISLKSIGISLGSPFHILTPDOEKPF 2160
Db 2101 WKAIQEGANSIVSLHQAACAAACLSRQASDSDSIISLKSIGISLGSPFHILTPDOEKPF 2160
QY 2161 SNKGPRILKPEKSTLTETKIESEKGIKGGKVKYSLITGKVRNSSEISGOMKQLOAN 2220
Db 2161 SNKGPRILKPEKSTLTETKIESEKGIKGGKVKYSLITGKVRNSSEISGOMKQLOAN 2220
QY 2221 MPSISRGKTIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGTATTTPRGAKEPVKS 2280
Db 2221 MPSISRGKTIHIPGVNRNSSSTSPVSKGPPPLKTPASKSPSEGTATTTPRGAKEPVKS 2280
QY 2281 ELSVPARQTSQIIGSSKAPSRSGRSTTPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
Db 2281 ELSVPARQTSQIIGSSKAPSRSGRSTTPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
QY 2341 KLSQLPRTSPTASTKSSGSGKVSYSKTPSGRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQLPRTSPTASTKSSGSGKVSYSKTPSGRQMSQOQLTKQTGLSKNASSIPRSESASKG 2400
QY 2401 LNQMNGANGKVKELSRMSSTKSSGSESDRSRPRVLVROSTFIKEAPPTLRKLEESA 2460
Db 2401 LNQMNGANGKVKELSRMSSTKSSGSESDRSRPRVLVROSTFIKEAPPTLRKLEESA 2460
QY 2461 SFESLSPSRPASPTRSQATPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
Db 2461 SFESLSPSRPASPTRSQATPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
QY 2521 RPAKRHDIAHSHSESRLPINSRGTWKREHSHKSSSLPRVSTWRRTGSSSSILSASSSS 2580
Db 2521 RPAKRHDIAHSHSESRLPINSRGTWKREHSHKSSSLPRVSTWRRTGSSSSILSASSSS 2580
QY 2581 SEKAKSEDEKHNVSISGTSKQENQVSAKGTWRKIKENEFSPNTSQTSSGATNGASS 2640
Db 2581 SEKAKSEDEKHNVSISGTSKQENQVSAKGTWRKIKENEFSPNTSQTSSGATNGASS 2640
QY 2641 KTLIYQMAPAVKTEDVWVRIEDCPIINPRSGSPGTNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQMAPAVKTEDVWVRIEDCPIINPRSGSPGTNTPPVIDSVSEKANPNIKDSKN 2700
QY 2701 QAKQNVGNSVPMRTVGLNRLTSFIQVADAPQOKGTEIKPGQNNPVVSETNESPIVET 2760
Db 2701 QAKQNVGNSVPMRTVGLNRLTSFIQVADAPQOKGTEIKPGQNNPVVSETNESPIVET 2760
QY 2761 PFSSSSSKHSSPSGTVAARVTPFNTPNPRKSSADSTGARPSSQIPTPVNNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAARVTPFNTPNPRKSSADSTGARPSSQIPTPVNNNTKKRDSKT 2820
QY 2821 DSTESSGTSQSPKSHSGSLVTSV 2843

Db	2821	DSTESSGTQSPKRGHSGSYLTVTSV	2843	
RESULT 12				
AAAY70304				
ID	AAAY70304	standard; protein; 2973 AA.		
XX	AAAY70304;			
AC	06-JUN-2000	(first entry)		
XX		Protein used in cancer diagnosis associated with APC/beta-catenin.		
DE		Human Tcf-4E protein; transcription factor; beta-catenin;		
XX		Adenomatous polyposis coli; APC; transcriptional activation;		
KW		Tcf responsive reporter gene; APC transcriptional regulatory pathway;		
KW		familial adenomatous polyposis; FAP; cancer; colorectal; thyroid; brain;		
KW		medulloblastoma; breast; head; neck; desmoid tumour; osteoma; cytostatic.		
OS		Homo sapiens.		
XX		WO200011195-A1.		
PN		02-MAR-2000.		
XX		20-AUG-1999;	99WO-US018774.	
XX		20-AUG-1998;	98US-00136605.	
XX		(UWJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.		
XX		He T, Vogelstein B, Kinzler KW;		
PI		WPI; 2000-237657/20.		
XX		Determining wild-type adenomatous polyposis coli protein for diagnosing		
PT		cancer comprises introducing a Tcf responsive reporter gene having		
PT		upstream sequences of c-MYC into a cell.		
XX		Disclosure; Page 58-65; 70pp; English.		
XX		The present sequence is a protein involved in cancer diagnosis associated		
CC		with APC or beta-catenin mutations. Human Tcf-4E protein expressed in		
CC		colorectal epithelium transactivates transcription when associated with		
CC		beta-catenin. Adenomatous polyposis coli (APC) regulates this		
CC		transcriptional activation, at least in part by binding to beta-catenin.		
CC		Determining wild-type APC protein for diagnosing cancer comprises		
CC		introducing a Tcf responsive reporter gene having upstream sequences of c		
CC		-MYC into a cell and measuring transcription of the reporter gene. The		
CC		candidate drug identified is useful for treating familial adenomatous		
CC		polyposis patients with APC or beta-catenin mutations and patients with		
CC		increased risk of developing cancers such as colorectal, thyroid, brain,		
CC		medulloblastoma, desmoid tumour, osteoma, breast, head and neck		
XX		Sequence 2973 AA;		
SQ		Query Match 99.7%; Score 14526; DB 3; Length 2973;		
		Best Local Similarity 99.7%; Pred. No. 0;		
		Matches 2835; Conservative 2; Mismatches 6; Indels 0; Gaps 0;		
QY	1	MAAASYDQLLKQVEALKMENSRLRQELDNLNHLTKLETEASNKREVLKQLOGSIEDAM	60	
Db	1	MAAASYDQLLKQVEALKMENSRLRQELDNLNHLTKLETEASNKREVLKQLOGSIEDAM	60	
QY	61	ASSGQIDLLERLKEMLNLDNSNFGVKLRKMSLRSGVSGSRGSCSPVPMGSPPR	120	
Db	61	ASSGQIDLLERLKEMLNLDNSNFGVKLRKMSLRSGVSGSRGSCSPVPMGSPPR	120	
QY	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKDWWYQAQLNLTKRIDSPLTENFSL	180	
Db	121	RGFVNGSRSTGYLEELKERSLLADLDKEEKDWWYQAQLNLTKRIDSPLTENFSL	180	
QY	181	QTDLTRQLEEARQIRVANEEOQLGTCQDMKRAQRRIARIQOIEKDIIRQLQSQAT	240	

Db	181	QTDLTRQLEEARQIRVANEEOQLGTCQDMKRAQRRIARIQOIEKDIIRQLQSQAT	240	
QY	241	BAERSSQNKHETGSHDAERQNEGQGVGEINMATSGQSGSTRMDHETASVLSSTHSA	300	
Db	241	BAERSSQNKHETGSHDAERQNEGQGVGEINMATSGQSGSTRMDHETASVLSSTHSA	300	
QY	301	PRRLTSHLGTQVEMVYSLLSMLGTHDKDMSTILLAMSSQSDSCISMRSQGLPLLIQL	360	
Db	301	PRRLTSHLGTQVEMVYSLLSMLGTHDKDMSTILLAMSSQSDSCISMRSQGLPLLIQL	360	
QY	361	HGNDKDSVLLGNSRGSKEARARASAAALHNIHQSOPDKRGREIRVHLHLEQIRAYCETC	420	
Db	361	HGNDKDSVLLGNSRGSKEARARASAAALHNIHQSOPDKRGREIRVHLHLEQIRAYCETC	420	
QY	421	WEQQAHEFGMDQDNMPAPVEHQICPAVCVLMKLSFDBEHRHAMELGGLOAIELLO	480	
Db	421	WEQQAHEFGMDQDNMPAPVEHQICPAVCVLMKLSFDBEHRHAMELGGLOAIELLO	480	
QY	481	VCEMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWALVAQKSESEDL	540	
Db	481	VCEMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWALVAQKSESEDL	540	
QY	541	QQVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSVLASLWLSAHT	600	
Db	541	QQVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSVLASLWLSAHT	600	
QY	601	ENKADICAVDGAFLVGLTYRSQNTLTAIESGGIILRNVSLLIATNEDHQILRENN	660	
Db	601	ENKADICAVDGAFLVGLTYRSQNTLTAIESGGIILRNVSLLIATNEDHQILRENN	660	
QY	661	CLQTLQHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLNLIHSHKMIAM	720	
Db	661	CLQTLQHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSMKLNLIHSHKMIAM	720	
QY	721	GSAALRNLMANPAKYKDANIMSPGSLPSLHVLRKQKALEAELDAHLSEFDNIDNLS	780	
Db	721	GSAALRNLMANPAKYKDANIMSPGSLPSLHVLRKQKALEAELDAHLSEFDNIDNLS	780	
QY	781	PKASHRSKORHKOSLYGDYVFDNTRDDNRSNFTGNMTVLSPLYNTTLVPSSSSSRGS	840	
Db	781	PKASHRSKORHKOSLYGDYVFDNTRDDNRSNFTGNMTVLSPLYNTTLVPSSSSSRGS	840	
QY	841	LDSSREKDSLRERIGILGNVHPATENPGTSKRGLOISITTAQAIKAWMEVSAIHTS	900	
Db	841	LDSSREKDSLRERIGILGNVHPATENPGTSKRGLOISITTAQAIKAWMEVSAIHTS	900	
QY	901	QEDRSSGTTTELHCVTDERNALRRSAAHSTNYNFTKSENNRCTCSMPYAKLEYKRSS	960	
Db	901	QEDRSSGTTTELHCVTDERNALRRSAAHSTNYNFTKSENNRCTCSMPYAKLEYKRSS	960	
QY	961	NDLSNVSNSNDGYGKRGQMKPSIESYSEDDSKFCSYGQVPADLAHKIHSANHMDNDGE	1020	
Db	961	NDLSNVSNSNDGYGKRGQMKPSIESYSEDDSKFCSYGQVPADLAHKIHSANHMDNDGE	1020	
QY	1021	LDTPINYSILKYSDHOLNSGRQSPQNERWARPKHITTEIKQSEQRORNSQSTTYPVTE	1080	
Db	1021	LDTPINYSILKYSDHOLNSGRQSPQNERWARPKHITTEIKQSEQRORNSQSTTYPVTE	1080	
QY	1081	STDDKHLKFPQHFQOQCVSPYRSRGANGSETNRVSGNHGINQNVQSGLCOEDDYEDDKP	1140	
Db	1081	STDDKHLKFPQHFQOQCVSPYRSRGANGSETNRVSGNHGINQNVQSGLCOEDDYEDDKP	1140	
QY	1141	TNYSERYSEBEOHEEERPTNYSIKYNEEKGHVDPIDYSLKYATDTPSSQKQSFSPKS	1200	
Db	1141	TNYSERYSEBEOHEEERPTNYSIKYNEEKGHVDPIDYSLKYATDTPSSQKQSFSPKS	1200	
QY	1201	SSGQSSKTEHMSSESSENTSTPSSNAKQONOLHPSAQRSGQPOKAAATCKVSSINQETIQ	1260	
Db	1201	SSGQSSKTEHMSSESSENTSTPSSNAKQONOLHPSAQRSGQPOKAAATCKVSSINQETIQ	1260	
QY	1261	TYCVEDTPICFSRCSLSLSSAEDETGQNTTQDASANTLQIAETIKGIGTSADDPV	1320	

Db 1261 TYCVEDTPICFRCSSLSLSAEDIEGNCOTTQADSAANTLOIAEIKETIGRSADDPV 1320
 QY 1321 SEVPAVSHPRTKSSRLQSSLSSESARHKAVEPPSGAKSPSKGAOTPKSPHYYOET 1380
 Db 1321 SEVPAVSHPRTKSSRLQSSLSSESARHKAVEPPSGAKSPSKGAOTPKSPHYYOET 1380
 QY 1381 PLMFSTRCTSVSSLSDFSRSIASSVQSEPCSGMWGIIISPSDLPDSPGQTMPPSRSKTTP 1440
 Db 1381 PLMFSTRCTSVSSLSDFSRSIASSVQSEPCSGMWGIIISPSDLPDSPGQTMPPSRSKTTP 1440
 QY 1441 PPPOTACTKEVPKNAKTAKEKREGKQAAVNAVORVQLPDADTLHLFAFESTPDGF 1500
 Db 1441 PPPOTACTKEVPKNAKTAKEKREGKQAAVNAVORVQLPDADTLHLFAFESTPDGF 1500
 QY 1501 SCSSLSALSIDEPFIQKVELRIMPPVQENDNGNETESQPKESNENQKEAEKTIIDSE 1560
 Db 1501 SCSSLSALSIDEPFIQKVELRIMPPVQENDNGNETESQPKESNENQKEAEKTIIDSE 1560
 QY 1561 KOLLDDDDDDIELEECIIISAMPTKSRKGGKPAQTASKLPPPVARKSQLPVYKLLPS 1620
 Db 1561 KOLLDDDDDDIELEECIIISAMPTKSRKGGKPAQTASKLPPPVARKSQLPVYKLLPS 1620
 QY 1621 QNRLOPKHVSFTPGDDMPVYCEVETPIINFSTATSLDITIESPPNELAAGBVGGAQ 1680
 Db 1621 QNRLOPKHVSFTPGDDMPVYCEVETPIINFSTATSLDITIESPPNELAAGBVGGAQ 1680
 QY 1681 SGEFEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAECSAMPKGSHP 1740
 Db 1681 SGEFEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAECSAMPKGSHP 1740
 QY 1741 FRVKIMDQVQOASASSAENKQOLDKKKTPSPVKPIPOKTEYETRVKNAADSKNIN 1800
 Db 1741 FRVKIMDQVQOASASSAENKQOLDKKKTPSPVKPIPOKTEYETRVKNAADSKNIN 1800
 QY 1801 AERFSDNCKSKQNLKNSKDFNDKLPNNEDVRGSAFDPSPHYTPIEGTPYCFSRND 1860
 Db 1801 AERFSDNCKSKQNLKNSKDFNDKLPNNEDVRGSAFDPSPHYTPIEGTPYCFSRND 1860
 QY 1861 SLSSLPDDDDDVLSREKAEKRAKENKESEAKVTSHTLTNSQOQANKTQIAKQPINR 1920
 Db 1861 SLSSLPDDDDDVLSREKAEKRAKENKESEAKVTSHTLTNSQOQANKTQIAKQPINR 1920
 QY 1921 GQPKPILOKQSTFPQSSKDIPOKGAATDEKLQNFALENTPVCFSHNSLSLSDIDQENN 1980
 Db 1921 GQPKPILOKQSTFPQSSKDIPOKGAATDEKLQNFALENTPVCFSHNSLSLSDIDQENN 1980
 QY 1981 NKENEDIKETEPPDQSGEPKQASGYAPKSFHVEDTPVCFSRNSLSLSDIDSEDDLLQ 2040
 Db 1981 NKENEDIKETEPPDQSGEPKQASGYAPKSFHVEDTPVCFSRNSLSLSDIDSEDDLLQ 2040
 QY 2041 ECISAMPKKKPSRLKGNKXSPNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKKKPSRLKGNKXSPNMGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
 QY 2101 WKAIQEGANSIVSLHQAACLSRQASDSDSILSLKSGISLGSFHLTPDQEKPT 2160
 Db 2101 WKAIQEGANSIVSLHQAACLSRQASDSDSILSLKSGISLGSFHLTPDQEKPT 2160
 QY 2161 SNKGPRILPGEKSTLETKEIESEKGIKGGKVKYSLITGKVRNSIISGOMQKPLQAN 2220
 Db 2161 SNKGPRILPGEKSTLETKEIESEKGIKGGKVKYSLITGKVRNSIISGOMQKPLQAN 2220
 QY 2221 MPSISGRRTMIHPGVNRSSSTSPVSKGPPKTPASKSPSEGQTATTPRGAKPSVKS 2280
 Db 2221 MPSISGRRTMIHPGVNRSSSTSPVSKGPPKTPASKSPSEGQTATTPRGAKPSVKS 2280
 QY 2281 ELSPVARQTSQIGGSKAPSRSGSDSTSPRAQOPLSRPTQSPGRNSISGRNGISPPN 2340
 Db 2281 ELSPVARQTSQIGGSKAPSRSGSDSTSPRAQOPLSRPTQSPGRNSISGRNGISPPN 2340
 QY 2341 KLSQPLRTSSPTASTKSGSGKMSYTFGRQMSQONLTQTGLSKNASSIPRESASKG 2400
 Db 2341 KLSQPLRTSSPTASTKSGSGKMSYTFGRQMSQONLTQTGLSKNASSIPRESASKG 2400

QY 2401 INQMNNGANKKVELSRMSTKSGSESDSRPVLVRQSTFIKEAPSPTLRRKLEESA 2460
 Db 2401 INQMNNGANKKVELSRMSTKSGSESDSRPVLVRQSTFIKEAPSPTLRRKLEESA 2460
 QY 2461 SFESLSPSRRPASPTRSOAQTPVLSPLPDMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 Db 2461 SFESLSPSRRPASPTRSOAQTPVLSPLPDMSLSTHSSVQAGGWRKLPPLNLSPTIEYNDG 2520
 QY 2521 RPAKHDIARSHSESPSRPLPINRSGTWKREHSEKSSSLPRVSTWRRTCSSSSILSASSES 2580
 Db 2521 RPAKHDIARSHSESPSRPLPINRSGTWKREHSEKSSSLPRVSTWRRTCSSSSILSASSES 2580
 QY 2581 SEKAKSEDEKHVNSISGTKQSKENOVSAKGTWRKIKENEFSPNTSTQTVSSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSISGTKQSKENOVSAKGTWRKIKENEFSPNTSTQTVSSSGATNGAES 2640
 QY 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTTPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTTPVIDSVSEKANPNIKDSKN 2700
 QY 2701 QAKQNVGSGVPMETVGLNRLTSFIOVDAPDQKGTETIKPGQNNPVPVSETNESPVERT 2760
 Db 2701 QAKQNVGSGVPMETVGLNRLTSFIOVDAPDQKGTETIKPGQNNPVPVSETNESPVERT 2760
 QY 2761 PFSSSSSKHSSPSGTVAAARVTPENYNPSPRKGSADSTARSQIPTPVNNTKKRDSKT 2820
 Db 2761 PFSSSSSKHSSPSGTVAAARVTPENYNPSPRKGSADSTARSQIPTPVNNTKKRDSKT 2820
 QY 2821 DSTESSGTQSPRHSGSYLVTSV 2843
 Db 2821 DSTESSGTQSPRHSGSYLVTSV 2843

RESULT 13

AAR58634
 ID AAR58634 standard; protein; 2843 AA.

XX AAR58634;

DT 25-MAR-2003 (revised)
 DT 21-JUN-1995 (first entry)

XX Adenomatous polyposis coli protein (APC).

KW Adenomatous polyposis coli protein gene; familial adenomatous polyposis;
 KW colorectal tumor; adenoma.

XX Homo sapiens.

XX W09421814-A1.

XX 29-SEP-1994.

XX 21-MAR-1994; 94WO-US002987.

XX 19-MAR-1993; 93US-00034850.

XX (UYJO) UNIV JOHNS HOPKINS.

XX Hall DE, Johnson KA, Kinzler KW, Vogelstein B;

XX WPI; 1994-317033/39.

XX N-PSDB; AAQ70633.

PT Antibodies to adenomatous polyposis coli protein - are used for detecting
 PT mutations in the APC gene for predicting pre-disposition to cancer,
 PT partic. colon cancer.

XX Claim 1; Page 46; 81pp; English.

XX Antibodies prepared to this protein are used for detecting mutations in
 CC the APC gene for predicting pre-disposition to cancer, particularly colon

CC cancer. The Abs can detect germ line or somatic mutations indicating a
CC predisposition to colon cancer and possibly gastric, oesophageal,
CC pancreatic or small cell lung cancers. (Updated on 25-MAR-2003 to correct
CC PN field.)

Sequence 2843 AA;

Query Match 99.6%; Score 14521; DB 2; Length 2843;

Query Match 99.8%; Score 14327;
Best Local Similarity 99.6%; Pred. No. 0;

BEST LOCAL SIMILARITY	99.0%;	FREQ. NO. 0;
Matches 2833;	Conservative	6; Mismatches
		4; Indels
		0; Gaps
		0;

Qy	1	MAAASYDOLLKQVPAKXWNSNLQOELEDNSNHLTKLETEASNKWEVLKOLQGSIEDPAM	60
Db	1	MAAASYDOLLKQVEALKXWNSNLQOELEDNSNHLTKLETEASNKWEVLKOLQGSIEDAI	60
Qy	61	ASSGOIDLLBELKELNLDSSNFPQVKLRSKMSLRASYGSRGVSRRSGECSVPWGMGFPR	120
Db	61	ASSGOIDLLBELKELNLDSSNFPQVKLRSKVSLASYGSRGVSRRSGECSVPWGMGFPR	120
Qy	121	RGFVNGSRESTGYLELEKERSLLADLDEKEKOWYIAOLQNLTKRIDSPLPTENPSL	180
Db	121	RGFVNGSRESTGYLELEKERSLLADLDEKEKDWYIAOLQNLTKRIDSPLPTENPSL	180
Qy	181	QTDLTRQLEVEARQIRVAMEEQIGTQDMEKRAORRIARIQOIEKDILRIRQLLOSOAT	240
Db	181	QTDLTRQLEVEARQIRVAMEEQIGTQDMEKRAORRIARIQOIEKDILRIRQLLOSOAT	240
Qy	241	EAERSSQNKHETGSHDAERQNEGQVGEINMATSGOGSSTRMDHETASVLSSSSTHSA	300
Db	241	EAERSSQNKHETGSHDAERQNEGQVGEINMATSGOGSSTRMDHETASVLSSSSTHSA	300
Qy	301	PRRLTSHLGTQVEMYISLLSMLGTHDKDDMSERTILLAMSSQDSCISMEQSCPLPILLOLL	360
Db	301	PRRLTSHLGTQVEMYISLLSMLGTHDKDDMSERTILLAMSSQDSCISMEQSCPLPILLOLL	360
Qy	361	HGNDKDSVLLGNRSRGSKEARASAAALHNI1HSQPDCKRGREIRVLHLLQIRAYCECTC	420
Db	361	HGNDKDSVLLGNRSRGSKEARASAAALHNI1HSQPDCKRGREIRVLHLLQIRAYCECTC	420
Qy	421	WEMQEAHEPGMDQDNMPAPVHEHQICPAVCVLMKLSFDEBHRHAMELGGLOIAEALLQ	480
Db	421	WEMQEAHEPGMDQDNMPAPVHEHQICPAVCVLMKLSFDEBHRHAMELGGLOIAEALLQ	480
Qy	481	VDCEMYGLTNDHYSITILRYAGMALTNLTFGDVANKATLCSMKGCWRALVAQLKSEBDL	540
Db	481	VDCEMYGLTNDHYSITILRYARMAALTNLTFGDVANKATLCSMKGCWRALVAQLKSEBDL	540
Qy	541	QOVTASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSVLASLWNL5AHCT	600
Db	541	QOVTASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSVLASLWNL5AHCT	600
Qy	601	ENKADICADVAGALAFVLGTLTVRSQNTNLAI1TESGGGILRNVS5LIATNEDHQI1RENN	660
Db	601	ENKADICADVAGALAFVLGTLTVRSQNTNLAI1TESGGGILRNVS5LIATNEDHQI1RENN	660
Qy	661	CIQTLLOHLKSHSLITVSNACCTGLNLSARNPKDQBALMDGAVSMKLNLIH5KHKNIAM	720
Db	661	CIQTLLOHLKSHSLITVSNACCTGLNLSARNPKDQBALMDGAVSMKLNLIH5KHKNIAM	720
Qy	721	GSAAALRNLMANRPKAYKXANIMSPGSSLP5LHVRKQKALEAELDAQH1SETFDNIDNLS	780
Db	721	GSAAALRNLMANRPKAYKXANIMSPGSSLP5LHVRKQKALEAELDAQH1SETFDNIDNLS	780
Qy	781	PKASHRSKORHKQSLGYDVYFDTRNDRDDNRDNFNFTGNMTVLSPLYLNTVTPLSSSSSRGS	840
Db	781	PKASHRSKORHKQSLGYDVYFDTRNDRDDNRDNFNFTGNMTVLSPLYLNTVTPLSSSSSRGS	840
Qy	841	LQSSRSEKORS1ERESGIGLGNVHPATEPNPCTSSKKGLOI1STTAAQIAKMWEEVSAIH7S	900
Db	841	LQSSRSEKORS1ERESGIGLGNVHPATEPNPCTSSKKGLOI1STTAAQIAKMWEEVSAIH7S	900
Qy	901	QEDRSSGSTTELCVTDERNALERSAAH7SHNTYNTFKSENSNRTCSPYAKLEFYKRS	960

Qy	2041	ECISSAMPKKKPSRLKGDNEKHSPRNMGGLIGEDLTLDLKDIOQPSEHGLSPDSNFD	2100
Db	2041	ECISSAMPKKKPSRLKGDNEKHSPRNMGGLIGEDLTLDLKDIOQPSEHGLSPDSNFD	2100
Qy	2101	WKAIQEGANSIVSSLHQAAAAACLSRQASDSDSILSKSGISILGSPFHLTPDQEEKPFT	2160
Db	2101	WKAIQEGANSIVSSLHQAAAAACLSRQASDSDSILSKSGISILGSPFHLTPDQEEKPFT	2160
Qy	2161	SNKGPRILKPGKEKSTLETTKIIESSEKGIKGGKKYKXSLITKGRVRSNEISGQWKQPLOAN	2220
Db	2161	SNKGPRILKPGKEKSTLETTKIIESSEKGIKGGKKYKXSLITKGRVRSNEISGQWKQPLOAN	2220
Qy	2221	MPSISRGRWTIHIPGVNRNSSSSTSPVSKKGPPLKTPASKSPSEGOATTTPRCAKPSVK	2280
Db	2221	MPSISRGRWTIHIPGVNRNSSSSTSPVSKKGPPLKTPASKSPSEGOATTTPRCAKPSVK	2280
Qy	2281	ELSPVARQTSOIGSSKAPRSRSGSRDTPRPAOQPLSRPIQSPGRNSISPGRNGISPPN	2340
Db	2281	ELSPVARQTSOIGSSKAPRSRSGSRDTPRPAOQPLSRPIQSPGRNSISPGRNGISPPN	2340
Qy	2341	KLSQLPRTSSPTASTKSSSGKMSYTSPGRQMSQQNLTKQTGLSKWASSIPRSESASKG	2400
Db	2341	KLSQLPRTSSPTASTKSSSGKMSYTSPGRQMSQQNLTKQTGLSKWASSIPRSESASKG	2400
Qy	2401	LNQMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA	2460
Db	2401	LNQMNNGANKKVELSRMSTKSSGESDRSRPVLVROSTFIKEAPSTLRKLEESA	2460
Qy	2461	SFESLSPSSRPASPTRSOATPVLSPSLPMWSLSTHSSVQAGWRKLPNLSPTIYNDG	2520
Db	2461	SFESLSPSSRPASPTRSOATPVLSPSLPMWSLSTHSSVQAGWRKLPNLSPTIYNDG	2520
Qy	2521	RPAKRHDIARSHGSPSRLPINRSGTWKREHSHSSSLPRVSTWRTTGGSSSILSASSS	2580
Db	2521	RPAKRHDIARSHGSPSRLPINRSGTWKREHSHSSSLPRVSTWRTTGGSSSILSASSS	2580
Qy	2581	SEKAKSEDEKHVNSISGTKOSKENQVSAKTWKIKENEPSPNISTQTVSSGATNGAES	2640
Db	2581	SEKAKSEDEKHVNSISGTKOSKENQVSAKTWKIKENEPSPNISTQTVSSGATNGAES	2640
Qy	2641	KTLIIYQAPAVSKTEDVWVRIEDCPINNPSSGRSPTCGTPPVIDSSEKANPNIKDSKN	2700
Db	2641	KTLIIYQAPAVSKTEDVWVRIEDCPINNPSSGRSPTCGTPPVIDSSEKANPNIKDSKN	2700
Qy	2701	QAKQNVGSGVPWRTVGLENRLTSFTQVADAPDQKTEIKPGQNPVPVSETNESPIVERT	2760
Db	2701	QAKQNVGSGVPWRTVGLENRLTSFTQVADAPDQKTEIKPGQNPVPVSETNESPIVERT	2760
Qy	2761	PFSSSSSSKHSSPSGTVAAARVTFPNYNPSPRKSADSTARSQIPTPVNNNTKKRDSKT	2820
Db	2761	PFSSSSSSKHSSPSGTVAAARVTFPNYNPSPRKSADSTARSQIPTPVNNNTKKRDSKT	2820
Qy	2821	DSTESSGTQSPKHHSGSYLTVSV	2843
Db	2821	DSTESSGTQSPKHHSGSYLTVSV	2843

DEC 11 1964

RESULT 14
A B C D E O A

AAR63508

ID AAR6

XX

AC AAR6

XX

DT 25-M

DT 23-M

XX
XX
TQ
TM-67DE
XXDE
Aden

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

KW Adenocarcinoma

KW fami:

KW
tumo:

XX

OS	Homo sapiens.
XX	
XX	US5352775-A.
XX	
XX	04-OCT-1994.
XX	
XX	08-AUG-1991; 91US-00741940.
XX	
XX	16-JAN-1991; 91GB-00000962.
PR	16-JAN-1991; 91GB-00000963.
PR	16-JAN-1991; 91GB-00000974.
PR	16-JAN-1991; 91GB-00000975.
XX	
XX	(ICIL) IMPERIAL CHEM IND.
PA	(CANC-) CANCER INST.
PA	(UYJO) UNIV JOHNS HOPKINS.
FA	(UTAH) UNIV UTAH.
XX	
XX	Markham AF, Hedge PJ, Anand R, Nakamura Y, Groden J, Kinzler K;
PI	Talliveris A, Carlson M, Vogelstein B, Albertsen H, White RL;
PI	Joslyn G;
XX	
XX	WPI; 1994-316233/39.
DR	N-PSDB; AAQ72297.
DR	
XX	
PT	New human adenomatous polyposis coli DNA encoding tumour repressor - and
PT	derived primers and probes for diagnosis, prognosis and treatment of
PT	cancer.
XX	
PS	Claim 3; Col 71-84; 113pp; English.
XX	
CC	AAQ72297 is a cDNA isolated from the human adenomatous polyposis coli
CC	(APC) gene, it encodes the tumour repressors described in AAR63507 and
CC	AAR63508. Determination of alterations in APC or its expression products
CC	can be used for the diagnosis and prognosis of cancer e.g. colorectal,
CC	lung and breast tumours; and for determining predisposition to certain
CC	cancers such as familial adenomatous polyposis (FAP) and Gardner's
CC	syndrome. The wild type APC gene (or a part of it) can be used
CC	therapeutically to restore gene function, while primers and probes
CC	derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to
CC	detect mutations. Also APC proteins or analogues can be administered to
CC	compensate for a defective gene, and epithelial cells, or transgenic
CC	animals carrying a mutated APC allele are useful for detecting
CC	therapeutic agents able to suppress tumorigenesis. (Updated on 25-MAR-
CC	2003 to correct PA field.)
XX	
XX	Sequence 2842 AA;
SQ	

DE	Adenomatous polyposis coli tumour repressor.
XX	
KW	Adenomatous polyposis coli; tumour repressor
KW	familial adenomatous polyposis; cancer diagn
KW	tumorigenesis suppression.

Db 300 PRRLTSHLGTKEVYVYSLMLGTHDKDDMRTRLILAMSSQDSCTSRFQSCCLPELLIQLL 359
QY 361 HGNDKDSVLLGNRSGSKARARASAAALHNIHISQPDCKRGRREIRVLHLLLOIRAYCETC 420
Db 360 HGNDKDSVLLGNRSGSKARARASAAALHNIHISQPDCKRGRREIRVLHLLLOIRAYCETC 419
QY 421 NEWQEAHFPGMDQDKNPNPAPVEHQICPAVCVLMKLSFDEHRHAMNBLGLQIAELLQ 480
Db 420 NEWQEAHFPGMDQDKNPNPAPVEHQICPAVCVLMKLSFDEHRHAMNBLGLQIAELLQ 479
QY 481 VDCENYGLTNDHYSITLARYAGMALTNLTFGDVANKATLCMSKCMRALVAQLKSEEDL 540
Db 480 VDCENYGLTNDHYSITLARYAGMALTNLTFGDVANKATLCMSKCMRALVAQLKSEEDL 539
QY 541 QOVIASVLRLNLSRADVNSKKTLEVGSVKALMECALEVKKESTLKSVLASALWNLASHCT 600
Db 540 QOVIASVLRLNLSRADVNSKKTLEVGSVKALMECALEVKKESTLKSVLASALWNLASHCT 599
QY 601 ENKADICAVDGAALFLVGLTYRSQNTNLAIIESGGILRNVSLLIATNEDHRQILRENN 660
Db 600 ENKADICAVDGAALFLVGLTYRSQNTNLAIIESGGILRNVSLLIATNEDHRQILRENN 659
QY 661 CLQTLLOHLKSHSLTVGNACGTLNLSARNPKDOEALWDMGAVSMLKNLIHSHKMIAM 720
Db 660 CLQTLLOHLKSHSLTVGNACGTLNLSARNPKDOEALWDMGAVSMLKNLIHSHKMIAM 719
QY 721 GSAAALRLMLNRPAPAKYKADANIMSGSSLPSLHVRYKQKALAEALDAHLSFTFONIDLS 780
Db 720 GSAAALRLMLNRPAPAKYKADANIMSGSSLPSLHVRYKQKALAEALDAHLSFTFONIDLS 779
QY 781 PKASHRSKORHKQSLYGYVFDPTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSRGSS 840
Db 780 PKASHRSKORHKQSLYGYVFDPTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSRGSS 839
QY 841 LDSSSEKDRSLERERGI GLGNYPATENPGTSSKGLQI8TTAAQIAKVMEEVSAIHTS 900
Db 840 LDSSSEKDRSLERERGI GLGNYPATENPGTSSKGLQI8TTAAQIAKVMEEVSAIHTS 899
QY 901 QEDRSSGTTLHCVTDERNALRRSSAAHTSNNTYNTFKTSENNTCSMPYAKLEYKSS 960
Db 900 QEDRSSGTTLHCVTDERNALRRSSAAHTSNNTYNTFKTSENNTCSMPYAKLEYKSS 959
QY 961 NDLSNSVSSNDGYKRGQMKPSIESYSEDDSKFCYGYQYFADLAHKIHSANHMDNDGE 1020
Db 960 NDLSNSVSSNDGYKRGQMKPSIESYSEDDSKFCYGYQYFADLAHKIHSANHMDNDGE 1019
QY 1021 LDTPIYSLKYDEQLNGRQSPQNERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1080
Db 1020 LDTPIYSLKYDEQLNGRQSPQNERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1079
QY 1081 STDDKHLKFPQHFQOECVSPYRSRGANGSENRVGSNHNQNVYQSLCOEDDYEDDKP 1140
Db 1080 STDDKHLKFPQHFQOECVSPYRSRGANGSENRVGSNHNQNVYQSLCOEDDYEDDKP 1139
QY 1141 TNYSEYSEEEQHEEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSPFSKSS 1200
Db 1140 TNYSEYSEEEQHEEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSPFSKSS 1199
QY 1201 SSGQSKTEHMSSESTSTPSSNAKRONLHPSSAQSRSGOPOKAACTCKYSSINOETIQ 1260
Db 1200 SSGQSKTEHMSSESTSTPSSNAKRONLHPSSAQSRSGOPOKAACTCKYSSINOETIQ 1259
QY 1261 TYCVEDTPICFRCSLSSLSAEDEIGCNQTOEADSAANTLQIAEIKGKIGTRSAEDPV 1320
Db 1260 TYCVEDTPICFRCSLSSLSAEDEIGCNQTOEADSAANTLQIAEIKGKIGTRSAEDPV 1319
QY 1321 SEVPAYVQHPRTKSRFLQSSLSSESARHAKAVEFFSGAKSPSKGAOCPKSPPEHYVQET 1380
Db 1320 SEVPAYVQHPRTKSRFLQSSLSSESARHAKAVEFFSGAKSPSKGAOCPKSPPEHYVQET 1379
QY 1381 PLMFSRCTSVSSLDSPESFESIASVQSEPCSGMVSGIISPSDLPDSPGOTMPPSRSKTTP 1440

Db 1380 PLMFSRCTSVSSLDSPESFESIASVQSEPCSGMVSGIISPSDLPDSPGOTMPPSRSKTTP 1439
QY 1441 PPQOTAQTKREVPKNKAPTAEKRESQPKQAAVNAAVQVRQVLPDADTLLHFAFESTPDGF 1500
Db 1440 PPQOTAQTKREVPKNKAPTAEKRESQPKQAAVNAAVQVRQVLPDADTLLHFAFESTPDGF 1499
QY 1501 SCSSSLSALSDBPFTQKOVLELRIMPVQENDNGNETESEQPKESNENQEKAEKTIIDSE 1560
Db 1500 SCSSSLSALSDBPFTQKOVLELRIMPVQENDNGNETESEQPKESNENQEKAEKTIIDSE 1559
QY 1561 KDLLDSDDDDDIIEBECIISAMPTKSSRKGGKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1560 KDLLDSDDDDDIIEBECIISAMPTKSSRKGGKPAQATASKLPPPVARKPSQLPVYKLLPS 1619
QY 1621 QNLQPKHVSFTPGDDMPRVYCEGTPIINFSTATSLDLTTIBSPNNELAAGVGREGGAQ 1680
Db 1620 QNLQPKHVSFTPGDDMPRVYCEGTPIINFSTATSLDLTTIBSPNNELAAGVGREGGAQ 1679
QY 1681 SGFEKRDITPTREGSTDEAOGGKTSVTIPELDDNKABEGDILACINCSAMPKGSHPK 1740
Db 1680 SGFEKRDITPTREGSTDEAOGGKTSVTIPELDDNKABEGDILACINCSAMPKGSHPK 1739
QY 1741 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1800
Db 1740 FRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1799
QY 1801 AERVFSDNKDSKKNLKNNSKDNKLPNNEDRVGSEAFDPSHHYTPTEGTTCFSRND 1860
Db 1800 AERVFSDNKDSKKNLKNNSKDNKLPNNEDRVGSEAFDPSHHYTPTEGTTCFSRND 1859
QY 1861 SLSSLPDDDDDVLSREKAEELRAKAKENKSEAKVTSHTELTSNQQSANKTQALAKQPINR 1920
Db 1860 SLSSLPDDDDDVLSREKAEELRAKAKENKSEAKVTSHTELTSNQQSANKTQALAKQPINR 1919
QY 1921 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLQNFATENTPVCFSHNSSSLSLSDIOENN 1980
Db 1920 GQPKPILOKQSTPQSSKDIIPDRGAATDEKLQNFATENTPVCFSHNSSSLSLSDIOENN 1979
QY 1981 NKNEPIKETEPDPSOQEPSPKQASGYAPKSFVEEDTPVCFSRNSLSLSIDSEDDLQ 2040
Db 1980 NKNEPIKETEPDPSOQEPSPKQASGYAPKSFVEEDTPVCFSRNSLSLSIDSEDDLQ 2039
QY 2041 ECISSAMPKKKPSRLKGDNEKHSRNMGGILGEDTLDKDQRPDSEHGLSPDSENF 2100
Db 2040 ECISSAMPKKKPSRLKGDNEKHSRNMGGILGEDTLDKDQRPDSEHGLSPDSENF 2099
QY 2101 WKAIQEGANSIVSLHQAACCLSCASSDPSILSLKSGISLGSPPHLLTPQEEKPFT 2160
Db 2100 WKAIQEGANSIVSLHQAACCLSCASSDPSILSLKSGISLGSPPHLLTPQEEKPFT 2159
QY 2161 SNKGPRILKPGEKSTLETTKKIESESGIKGKKVYKSLITGKVRNSSEISGQMKQPLQAN 2220
Db 2160 SNKGPRILKPGEKSTLETTKKIESESGIKGKKVYKSLITGKVRNSSEISGQMKQPLQAN 2219
QY 2221 MPSISRGRTMIHITPGVNSSSSTSPVSKGPPILKTASKSPSEGTATTSIPRGAKEBVS 2280
Db 2220 MPSISRGRTMIHITPGVNSSSSTSPVSKGPPILKTASKSPSEGTATTSIPRGAKEBVS 2279
QY 2281 ELSVPARQTSQIIGSSKAPSRSGRSTPSRPAQQLSRPIQSPGRNISIFGRNGISPPN 2340
Db 2280 ELSVPARQTSQIIGSSKAPSRSGRSTPSRPAQQLSRPIQSPGRNISIFGRNGISPPN 2339
QY 2341 KLSQLPRTSPSTASTKSGSGKMSYTPSORQMSQQNLTKQTGLSKVASSIPRSESASKG 2400
Db 2340 KLSQLPRTSPSTASTKSGSGKMSYTPSORQMSQQNLTKQTGLSKVASSIPRSESASKG 2399
QY 2401 LQNMWNGGANKVKVELSRMSSTKSSGESDRSRPVLVROSTTIKEAPSTPLRKLKLESA 2460
Db 2400 LQNMWNGGANKVKVELSRMSSTKSSGESDRSRPVLVROSTTIKEAPSTPLRKLKLESA 2459
QY 2461 SFESLSPSRPASPTRSQATPVLSPSLPDMVSLSTHSSVOAGGWRKLPPLNLSPTIEYNDG 2520
Db 2460 SFESLSPSRPASPTRSQATPVLSPSLPDMVSLSTHSSVOAGGWRKLPPLNLSPTIEYNDG 2519

QY 2521 RPAKHDIARSHSPRLINSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASSPS 2580
 DB 2520 RPAKHDIARSHSPRLINSGTWKREHSHSSSLPRVSTWRRTGSSSSILSASSPS 2579
 QY 2581 SEKAKSEDEKHVNSISGTSKQKNOVSAKGTWRKIKENEPSPNTSQTVSSGATNGAES 2640
 DB 2580 SEKAKSEDEKHVNSISGTSKQKNOVSAKGTWRKIKENEPSPNTSQTVSSGATNGAES 2639
 QY 2641 KTLIIQMAPAVSKTDDVWRIEDCPINPRSGRSPGTNPVSDVSEKAPNIKDSKON 2700
 DB 2640 KTLIIQMAPAVSKTDDVWRIEDCPINPRSGRSPGTNPVSDVSEKAPNIKDSKON 2699
 QY 2701 QAKQNVGNGSVPMRTVGLNRLTSFIOVDAPDOKGTEIKPQNNPVVSETNESPIVERT 2760
 DB 2700 QAKQNVGNGSVPMRTVGLNRLTSFIOVDAPDOKGTEIKPQNNPVVSETNESPIVERT 2759
 QY 2761 PFSSSSSKHSSPSGTVAAVTPFNPNPSPRKSSADSTSARPSQIPTPVNNNTKKRDSKT 2820
 DB 2760 PFSSSSSKHSSPSGTVAAVTPFNPNPSPRKSSADSTSARPSQIPTPVNNNTKKRDSKT 2819
 QY 2821 DSTESSGTQPKRHSGSYLTVSV 2843
 DB 2820 DSTESSGTQPKRHSGSYLTVSV 2842

RESULT 15
 ABG90968
 ID ABG90968 standard; peptide; 2842 AA.
 XX AC ABG90968;
 XX DT 29-NOV-2002 (first entry)
 XX DE Human APC protein.
 XX KW Adenomatous polyposis coli; APC; human; neoplastic tissue;
 XX KW mutation detection; tumour; cancer.
 XX OS Homo sapiens.
 XX PN US6413727-B1.
 XX PD 02-JUL-2002.
 XX PF 25-MAY-1995; 95US-00449731.
 XX PR 16-JAN-1991; 91GB-00000962.
 XX PR 16-JAN-1991; 91GB-00000963.
 XX PR 16-JAN-1991; 91GB-00000974.
 XX PR 16-JAN-1991; 91GB-00000975.
 XX PR 08-AUG-1991; 91US-00741940.
 XX PR 12-AUG-1994; 94US-00289548.
 XX (UJGO) UNIV JOHNS HOPKINS.
 XX (UTAH) UNIV UTAH.
 XX (NICA-) JAPANESE FOUND CANCER RES.
 XX (ZENE) ZENECA LTD.
 XX Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ, Joslyn G;
 PI Kinzler K, Markham AF, Nakamura Y, Thliveris A, Vogelstein B;
 PI White RL;
 XX WPI; 2002-641559/69.
 XX Method to aid in the diagnosis/prognosis of neoplastic tissues in humans,
 PT by detecting somatic alteration of wild-type APC protein in tumor tissue
 PT isolated from human, the alteration indicating neoplasia of the tissue.
 XX Example 15; Fig 3; 140pp; English.
 XX This invention relates to a novel method to aid in the diagnosis or
 CC prognosis of a neoplastic tissue of a human. The method involves

CC detecting somatic alteration of wild-type adenomatous polyposis coli)
 CC protein in a tumour tissue isolated from a human (the alteration is
 CC indicating neoplasia of the tissue). The method of the invention is
 CC useful in diagnosis or prognosis of a neoplastic tissue of a human. The
 CC method is useful in detection of genetic predisposition to cancer. The
 CC present sequence represents a peptide sequence used in the method of the
 CC invention
 XX Sequence 2842 AA;
 SQ
 Query Match 99.6%; Score 14515.5; DB 5; Length 2842;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2835; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
 QY 1 MAASDYDQLLKQVAKVENSINLRQELSDNSHLTKLETSANNKEVYLKOLQGSIEDEAM 60
 DB 1 MAASDYDQLLKQVAKVENSINLRQELSDNSHLTKLETSANNKEVYLKOLQGSIEDEAM 60
 QY 61 ASSQOIDLLERLKELNLDSSNFPQVGLRSKMSLSYSGREGSVSSRGCECPVPMGSFPR 120
 DB 61 ASSQOIDLLERLKELNLDSSNFPQVGLRSKMSLSYSGREGSVSSRGCECPVPMGSFPR 120
 QY 121 RGFVNGSRESTGYLEELKERSLLADLDKEEKQWYVLAQLQNLTKRIDSLPTENFSL 180
 DB 121 RGFVNGSRESTGYLEELKERSLLADLDKEEKQWYVLAQLQNLTKRIDSLPTENFSL 179
 QY 181 QTDLTRQLEYEARQIRVAMEEQGLTQDMKRAQRRIARIQIEKDILRIQLQSQAT 240
 DB 180 QTDLTRQLEYEARQIRVAMEEQGLTQDMKRAQRRIARIQIEKDILRIQLQSQAT 239
 QY 241 EAERSSQNKHETGSHDAERQNEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 300
 DB 240 EAERSSQNKHETGSHDAERQNEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 299
 QY 301 PRRLTSHLGTVMVYISLLSMGLTHDXDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL 360
 DB 300 PRRLTSHLGTVMVYISLLSMGLTHDXDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL 359
 QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKRRREIRVHLLEQIRAYCETC 420
 DB 360 HGNDKDSVLLGNSRGSKEARASAAALHNIHSOPDDKRRREIRVHLLEQIRAYCETC 419
 QY 421 WEQEAHEPQMDQKNPMPAPVEHOICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 480
 DB 420 WEQEAHEPQMDQKNPMPAPVEHOICPAVCVLMKLSFDEEHRHAMNELGLOAIAELLQ 479
 QY 481 VDCMYGLTNDHYSITLRYAGMALTNLTGADVANKATLCMRKGMALVAQLKSESEDL 540
 DB 480 VDCMYGLTNDHYSITLRYAGMALTNLTGADVANKATLCMRKGMALVAQLKSESEDL 539
 QY 541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVLGALNLSAHC 600
 DB 540 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVLGALNLSAHC 599
 QY 601 ENKADICAVDQALAFVLGTLTYRSQTNLTALIESGGGILRNVSLSLIATNEDHRIQLENN 660
 DB 600 ENKADICAVDQALAFVLGTLTYRSQTNLTALIESGGGILRNVSLSLIATNEDHRIQLENN 659
 QY 661 CLQTLLOHLKSHSLTIYSNACGLTNLSARNPKDQALWDMGAVSMKLNLSHGHKMIAM 720
 DB 660 CLQTLLOHLKSHSLTIYSNACGLTNLSARNPKDQALWDMGAVSMKLNLSHGHKMIAM 719
 QY 721 GSAALRNLMANRPKYKDNINMSPGSSLPVSHVRKQKALEAEADQHLSETFDNIDNLS 780
 DB 720 GSAALRNLMANRPKYKDNINMSPGSSLPVSHVRKQKALEAEADQHLSETFDNIDNLS 779
 QY 781 PKASHRSKQKHQSLGYDYPDTRNDHNDSDPNTGNTVLPYINTTVLPSSSSRG 840
 DB 780 PKASHRSKQKHQSLGYDYPDTRNDHNDSDPNTGNTVLPYINTTVLPSSSSRG 839
 QY 841 LDSRSKDRSLERERIGLGNVHPATENPCTSKRGLQISTTAAQIAKVMEVSAIHTS 900
 DB 840 LDSRSKDRSLERERIGLGNVHPATENPCTSKRGLQISTTAAQIAKVMEVSAIHTS 899

QY 2761 PFSSSSSKHSPSGTVAARVTPFNPNPSPKSSADSTSARPSQIPTPVNNNTKRRDSKT 2820
Db 2761 PFSSSSSKHSPSGTVAARVTPFNPNPSPKSSADSTSARPSQIPTPVNNNTKRRDSKT 2820
QY 2821 DSESSGTSQPKRHSGSYLTVSV 2843
Db 2821 DSESSGTSQPKRHSGSYLTVSV 2843

RESULT 10
US-08-821-355A-7
; Sequence 7, Application US/08821355A
; Patent No. 5851775
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821.355A
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5851775e
US-08-821-355A-7

Query Match 99.7%; Score 14533; DB 2; Length 2973;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAASYDQLLKQVEALKMNSNLRQLEDNSNHLTKLETEASNMKVYLKQLQGSIEDAM 60
Db 1 MAAASYDQLLKQVEALKMNSNLRQLEDNSNHLTKLETEASNMKVYLKQLQGSIEDAM 60
QY 61 ASSGOIDLLERKELNLDSSNPPGVLKRSKMSLRSGREGSVSRSGECSPVPMGSPFR 120
Db 61 ASSGOIDLLERKELNLDSSNPPGVLKRSKMSLRSGREGSVSRSGECSPVPMGSPFR 120
QY 121 RGFVNGSRESTGYLBELEKERSLLLDLADLDBKEKEDWYQAQLNLTKRIDSPLTENFSL 180

121 RGFVNGSRESTGYLBELEKERSLLLDLADLDBKEKEDWYQAQLNLTKRIDSPLTENFSL 180
181 QDTLTRQLYEYARQIRVAMEEQLGTCODMEKAKAQRRIARIQOIEKDILIRIQLLOSQAT 240
181 QDTLTRQLYEYARQIRVAMEEQLGTCODMEKAKAQRRIARIQOIEKDILIRIQLLOSQAT 240
241 EAERSSONKHETGSHDAERQNEGGVGEINMATSNGQSGSTTTRMDHETASVLSSTHSA 300
241 EAERSSONKHETGSHDAERQNEGGVGEINMATSNGQSGSTTTRMDHETASVLSSTHSA 300
301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDDMSRLLLAWSQSDSCISMROSGCPLLIQLL 360
301 PRRLTSHLGTKEVMYVSLLSMLGTHDKDDMSRLLLAWSQSDSCISMROSGCPLLIQLL 360
361 HGNDKOSVLLGNRSGSKAPARASAAALHNIHSPQDDKRRRIIRVHLLEQIRAYCETC 420
361 HGNDKOSVLLGNRSGSKAPARASAAALHNIHSPQDDKRRRIIRVHLLEQIRAYCETC 420
421 WEQBEAHEPCMDODKNMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGLQAIJELQ 480
421 WEQBEAHEPCMDODKNMPAPVEHQICPAVCVLMKLSFDEEHRHANNELGGLQAIJELQ 480
481 VDCEMYGLTNDHYSITLRRYAGVALTNLTFDGVANKATLCSMKGCMRALVAQLKSESED 540
481 VDCEMYGLTNDHYSITLRRYAGVALTNLTFDGVANKATLCSMKGCMRALVAQLKSESED 540
541 QOVIASVLRNLSWRADVNSKTLREYGSVKALMECALEVUKKESTLKSVISALNLSAHT 600
541 QOVIASVLRNLSWRADVNSKTLREYGSVKALMECALEVUKKESTLKSVISALNLSAHT 600
601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILNRVSSLIATNEDHRRQILRENN 660
601 ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGILNRVSSLIATNEDHRRQILRENN 660
661 CLOTLLQHLKSHSLTIVSNACGTLNLSARNPKDQBALWDMGAVSMKLIHSHKHMIAM 720
661 CLOTLLQHLKSHSLTIVSNACGTLNLSARNPKDQBALWDMGAVSMKLIHSHKHMIAM 720
721 GSAALRNLMANRPAPYKDIANIMSPGSLPRLHVRKQKALEALDQHLSETFDNDLS 780
721 GSAALRNLMANRPAPYKDIANIMSPGSLPRLHVRKQKALEALDQHLSETFDNDLS 780
781 PKASHRSKQKHQSLYGDYVDTNRHDDNRSDNFTGNMTVLSPLYNTTLPSSSSSRGS 840
781 PKASHRSKQKHQSLYGDYVDTNRHDDNRSDNFTGNMTVLSPLYNTTLPSSSSSRGS 840
841 LDSRSSEKDSLRERGI GLGNYPATENPGTSKGLQISTTAAQIAKMEVSAIHTS 900
841 LDSRSSEKDSLRERGI GLGNYPATENPGTSKGLQISTTAAQIAKMEVSAIHTS 900
901 QEDRSSGTTTELHCVTDERNALRRSAAATHSNFTKSENSENRTCSMPYAKLEYKRSS 960
901 QEDRSSGTTTELHCVTDERNALRRSAAATHSNFTKSENSENRTCSMPYAKLEYKRSS 960
961 NDSLSVSSNDGYGKRGOMKPSIESYEDDESXFCYGYQYPADLAHKIHSANHMDNDGE 1020
961 NDSLSVSSNDGYGKRGOMKPSIESYEDDESXFCYGYQYPADLAHKIHSANHMDNDGE 1020
1021 LDTPIYNSLYSDEQLNSGRQSPQNERWAPKXIIIEIKQSEQRQSRNQSTTYPVTE 1080
1021 LDTPIYNSLYSDEQLNSGRQSPQNERWAPKXIIIEIKQSEQRQSRNQSTTYPVTE 1080
1081 STDKHLKLFQPHFGQECVSPYRSRGANGSETNRVGSNGHINQVNSQSLCQEDDYEDDKP 1140
1081 STDKHLKLFQPHFGQECVSPYRSRGANGSETNRVGSNGHINQVNSQSLCQEDDYEDDKP 1140
1141 TNSYERYSEEEQEEERPTNYSIKYNEKRVDPIDYSLKYATDIPSSQKOSFSPKS 1200
1141 TNSYERYSEEEQEEERPTNYSIKYNEKRVDPIDYSLKYATDIPSSQKOSFSPKS 1200
1201 SSQSSKTEHMSSSSENTSTPSSNAKRONQLHPSSAQSRSGQPKAATCKVSSINQETIQ 1260

Db 1201 SSGSSKTEHMSSENSTPSSNAKRONQLHPSSAQSGQKQKATCKVSSINOETIQ 1260
Qy 1261 TYCVEDTPICFRCSLSLSLSAEDEIGCQNTTQEADSANTLQIAIEIKGIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFRCSLSLSLSAEDEIGCQNTTQEADSANTLQIAIEIKGIGTRSAEDPV 1320
Qy 1321 SEVPAVSOHPRTKSSRLQGSLSSESARHAKAVFPSPGAKSPKSGAQTPKSPPEHYVQET 1380
Db 1321 SEVPAVSOHPRTKSSRLQGSLSSESARHAKAVFPSPGAKSPKSGAQTPKSPPEHYVQET 1380
Qy 1381 PLAFSRCTSVSSLDSPFSRGIASSVQPCSGMVSGIISPSDLPDGPQTMPPSRSKTPP 1440
Db 1381 PLAFSRCTSVSSLDSPFSRGIASSVQPCSGMVSGIISPSDLPDGPQTMPPSRSKTPP 1440
Qy 1441 PPQTAQTKREVKNKAPTAKEBESGPKQAANAAVQVQVLPDADTLHLHFAESTPDGF 1500
Db 1441 PPQTAQTKREVKNKAPTAKEBESGPKQAANAAVQVQVLPDADTLHLHFAESTPDGF 1500
Qy 1501 SCSSLSALSDEPFIQKVELRIMPVQNDNGNETESQPKESNENOEKAEKTIIDSE 1560
Db 1501 SCSSLSALSDEPFIQKVELRIMPVQNDNGNETESQPKESNENOEKAEKTIIDSE 1560
Qy 1561 KOLLDDSDDDIELEBCCIISAMPTKSSRGKPKPAQOTASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDIELEBCCIISAMPTKSSRGKPKPAQOTASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOPKHVSFTPGDDMPVYCVGPTPIINFSTATSLDLTIESPPNELAAGVGVRGGAQ 1680
Db 1621 QNRLOPKHVSFTPGDDMPVYCVGPTPIINFSTATSLDLTIESPPNELAAGVGVRGGAQ 1680
Qy 1681 SGEFEKDTIPTTEGRSTDEAGGKTSSVTIPELDDNKAEGDILAEICINSAMPKGSHPK 1740
Db 1681 SGEFEKDTIPTTEGRSTDEAGGKTSSVTIPELDDNKAEGDILAEICINSAMPKGSHPK 1740
Qy 1741 FRVKKINDVOQASASSAPKNQOLDGKKKTPSPYKPIPONTTEYTRVRKNADSKNNLN 1800
Db 1741 FRVKKINDVOQASASSAPKNQOLDGKKKTPSPYKPIPONTTEYTRVRKNADSKNNLN 1800
Qy 1801 AERFVSNKDSKQNLKNNKDNKLPNNEDVRGSAFSDSPHHVTPIEGTYPYCFSRND 1860
Db 1801 AERFVSNKDSKQNLKNNKDNKLPNNEDVRGSAFSDSPHHVTPIEGTYPYCFSRND 1860
Qy 1861 SLSLDFDDDDVLSREKAEIRKAKENKESAKVTSHTELTSQOQANKTQATKQPINR 1920
Db 1861 SLSLDFDDDDVLSREKAEIRKAKENKESAKVTSHTELTSQOQANKTQATKQPINR 1920
Qy 1921 GQPKPILQKOSTFPQSKDIPDRGAATDEKLQNFALIENTPVCFSHNSSLSLSLSDIOENN 1980
Db 1921 GQPKPILQKOSTFPQSKDIPDRGAATDEKLQNFALIENTPVCFSHNSSLSLSLSDIOENN 1980
Qy 1981 NKENEPIKETEPPDSQGEPSKQASGAPKSFHVEDTTPVCFSRNSSLSLSLSDIOENLLQ 2040
Db 1981 NKENEPIKETEPPDSQGEPSKQASGAPKSFHVEDTTPVCFSRNSSLSLSLSDIOENLLQ 2040
Qy 2041 ECISAMPKPKKBRLLKGDNEKHSRPNWGGILGLEDLTLDKIQRDSEHGLSPDSENF 2100
Db 2041 ECISAMPKPKKBRLLKGDNEKHSRPNWGGILGLEDLTLDKIQRDSEHGLSPDSENF 2100
Qy 2101 WKAIQEGANSIVSLHQAACCLSRQASDSDSILSKSGISLGSPFHLTPQOEKPF 2160
Db 2101 WKAIQEGANSIVSLHQAACCLSRQASDSDSILSKSGISLGSPFHLTPQOEKPF 2160
Qy 2161 SNKGPRILKKEKSTLTETKTESKGIKGGKVVYKSLITGKVRNSEISGQMKQPLQAN 2220
Db 2161 SNKGPRILKKEKSTLTETKTESKGIKGGKVVYKSLITGKVRNSEISGQMKQPLQAN 2220
Qy 2221 MPSISRGRTWTHIPGVNRNSSSTSPVSKGPPPLTKPASKSPSEGTATTSFPGAKPSVK 2280
Db 2221 MPSISRGRTWTHIPGVNRNSSSTSPVSKGPPPLTKPASKSPSEGTATTSFPGAKPSVK 2280
Qy 2281 ELSVARTQTSQIGSSKAPRSGSRDSTPRPAQOPLSRPIQSPGRNSISPGNGISPPN 2340
Db 2281 ELSVARTQTSQIGSSKAPRSGSRDSTPRPAQOPLSRPIQSPGRNSISPGNGISPPN 2340

Qy 2341 KLSQLPRTSPSTASTKSSSGSKMYSYTPGRMSQQNLTKOTGLSKVAASSIPRSESASK 2400
Db 2341 KLSQLPRTSPSTASTKSSSGSKMYSYTPGRMSQQNLTKOTGLSKVAASSIPRSESASK 2400
Qy 2401 LQNMNNGANKKVELSRMSSTKSSGSESDRSEBPVLVROSTFIKEAPSPTLARKLESA 2460
Db 2401 LQNMNNGANKKVELSRMSSTKSSGSESDRSEBPVLVROSTFIKEAPSPTLARKLESA 2460
Qy 2461 SFESLSPSRSPASPTRSQAOQTPVLSPSLPMSLSTHSSVOAGGWRKLPPLNLSPTIENDG 2520
Db 2461 SFESLSPSRSPASPTRSQAOQTPVLSPSLPMSLSTHSSVOAGGWRKLPPLNLSPTIENDG 2520
Qy 2521 RPAKHDIARSHSPSRPLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
Db 2521 RPAKHDIARSHSPSRPLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
Qy 2581 SEKAKSEDEKRVNISGTTQSKENQVSAKGTWRKIKENEFSPMTNSTSTQTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKRVNISGTTQSKENQVSAKGTWRKIKENEFSPMTNSTSTQTVSSGATNGAES 2640
Qy 2641 KTLIYQAPAVSKTEDVWRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
Db 2641 KTLIYQAPAVSKTEDVWRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKN 2700
Qy 2701 QAKQNVNGSVPMRTVGLNELRTSFIQVADPDQKGTBIKPGQNNPVPVSETNESPIVERT 2760
Db 2701 QAKQNVNGSVPMRTVGLNELRTSFIQVADPDQKGTBIKPGQNNPVPVSETNESPIVERT 2760
Qy 2761 PFSSSSSKHSPSGTVAARVTPFNYPNPSPKSSADSTARSPOIPTPVANNYKRDST 2820
Db 2761 PFSSSSSKHSPSGTVAARVTPFNYPNPSPKSSADSTARSPOIPTPVANNYKRDST 2820
Qy 2821 DSTESSGTCSPKHSYGLVTSV 2843
Db 2821 DSTESSGTCSPKHSYGLVTSV 2843

RESULT 11

US-09-003-687A-7

; Sequence 7, Application US/09003687A

; Patent No. 5998600

; GENERAL INFORMATION:

; APPLICANT: Barker, Nick

; APPLICANT: Clevers, Hans

; APPLICANT: Korinek, Vladimir

; APPLICANT: Morin, Patrice

; APPLICANT: Kinzler, Kenneth

; APPLICANT: Vogelstein, Bert

; APPLICANT: Sparks, Andrew

; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESS: Banner & Witcoff, Ltd.

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20001

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/003.687A

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/821,355

; FILING DATE: 20-MAR-1997

; ATTORNEY/AGENT INFORMATION:

```

; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5998600e
; US-09-003-687A-7

Query Match          99.7%; Score 14533; DB 2; Length 2973;
Best Local Similarity 99.8%; Pred No. 0;
Matches 2936; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAASYDQLLKQVEALKMENSNLRLQLEDNSNHLTKLETEASNKMKVLLKQLQGSIEDEAM 60
DB 1 MAAASYDQLLKQVEALKMENSNLRLQLEDNSNHLTKLETEASNKMKVLLKQLQGSIEDEAM 60
QY 61 ASSGOIDLLEKLKELNLDSSNPPGVKLRKMSLSYSGSREGSVSRSGECSVPWMSGPPR 120
DB 61 ASSGOIDLLEKLKELNLDSSNPPGVKLRKMSLSYSGSREGSVSRSGECSVPWMSGPPR 120
QY 121 RGFVNGSRSTGYLLEEKERSLLIADLDKEEKEDWYIAQLNLTKRIDSPLTENFSL 180
DB 121 RGFVNGSRSTGYLLEEKERSLLIADLDKEEKEDWYIAQLNLTKRIDSPLTENFSL 180
QY 181 QTDLTROLEYEARQIRVAMEZQLGTCQDMKRAQRRIARIQOIEKOLIRIQLLOQAT 240
DB 181 QTDLTROLEYEARQIRVAMEZQLGTCQDMKRAQRRIARIQOIEKOLIRIQLLOQAT 240
QY 241 EAEASSQNKHTGTSHDAERQEGQGVGEINNATSGNGGCGSTRMDHETASVLSSSSTHSA 300
DB 241 EAEASSQNKHTGTSHDAERQEGQGVGEINNATSGNGGCGSTRMDHETASVLSSSSTHSA 300
QY 301 PRRLTSHLGTQVEMVYSLLSMLGTHDKDDMSRTLLAMSSODSCISMRQSGCLPLLIQLL 360
DB 301 PRRLTSHLGTQVEMVYSLLSMLGTHDKDDMSRTLLAMSSODSCISMRQSGCLPLLIQLL 360
QY 361 HGNDKDSVLLNGSRGSKARASAAALHNIITHSQDDDKRGREIRVLHLLQIRAYCETC 420
DB 361 HGNDKDSVLLNGSRGSKARASAAALHNIITHSQDDDKRGREIRVLHLLQIRAYCETC 420
QY 421 WEWQEAHEPGMDQDQKNPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAELLQ 480
DB 421 WEWQEAHEPGMDQDQKNPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAELLQ 480
QY 481 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEBDL 540
DB 481 VDCMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEBDL 540
QY 541 QOVIASVLRLNSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSALHACT 600
DB 541 QOVIASVLRLNSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSALHACT 600
QY 601 ENKADICAVDGLALFLVGLTYRSTNTLAIIESGGGILRNWSSLIATNEDHROILRENN 660
DB 601 ENKADICAVDGLALFLVGLTYRSTNTLAIIESGGGILRNWSSLIATNEDHROILRENN 660
QY 661 CLQTLLOHLKSHSLTIVSNACOTLNLNLSARPDKQDALMDVAGVSMKLNLIHSHKMIAM 720
DB 661 CLQTLLOHLKSHSLTIVSNACOTLNLNLSARPDKQDALMDVAGVSMKLNLIHSHKMIAM 720
QY 721 GSAAALRLNMANRPYKADANIMSGSSLPLSLHVRKQKALEABDLAQHLSLTFDNILNS 780
DB 721 GSAAALRLNMANRPYKADANIMSGSSLPLSLHVRKQKALEABDLAQHLSLTFDNILNS 780
QY 781 PKASHRSKQRHKQSLYGDYVFDTNHDDNRSDNFNTGNMTVLSPLYNTTVLPSSSSSRGS 840

```

```

781 PKASHRSKQRHKQSLYGDYVFDTNHDDNRSDNFNTGNMTVLSPLYNTTVLPSSSSSRGS 840
841 LDSRSSEKORSLEBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS 900
841 LDSRSSEKORSLEBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS 900
901 QEDRSSGSGTTELHCVTDERNALRRSSAAATHGNTYNTFKSENSNRCTCSMPYAKLEYKRS 960
901 QEDRSSGSGTTELHCVTDERNALRRSSAAATHGNTYNTFKSENSNRCTCSMPYAKLEYKRS 960
961 NDSLSNVSSNDGYGKQGMKPSIESYSEDDSKFCYGOYPADLAHIIHSAHMDNDGE 1020
961 NDSLSNVSSNDGYGKQGMKPSIESYSEDDSKFCYGOYPADLAHIIHSAHMDNDGE 1020
1021 LDTPIYSLKYSDQNLNCRQSPQONERWARPKHIIEDIKQSEQRQSNQSTTYPVYTE 1080
1021 LDTPIYSLKYSDQNLNCRQSPQONERWARPKHIIEDIKQSEQRQSNQSTTYPVYTE 1080
1081 STDDKHLKFPQHPGQECVSPYRSGANGSETRVNSNGHINQNVSQSLCQEDDDYEDDKP 1140
1081 STDDKHLKFPQHPGQECVSPYRSGANGSETRVNSNGHINQNVSQSLCQEDDDYEDDKP 1140
1141 TNYSEYSEEEQHEEERPTNYSIKYNEKRVHVDQPIDYSLKYATDIPSSQKQSPFSKS 1200
1141 TNYSEYSEEEQHEEERPTNYSIKYNEKRVHVDQPIDYSLKYATDIPSSQKQSPFSKS 1200
1201 SSGQSKTEHMSSESTSTPSNNAKQNLHPSSAQSRSGQPQKAATCKVSSINQETIQ 1260
1201 SSGQSKTEHMSSESTSTPSNNAKQNLHPSSAQSRSGQPQKAATCKVSSINQETIQ 1260
1261 TYCVDPTPCFSCSSLSLSAEDIGCNOTOEAASANTLOIAIEIKIGITRSADPV 1320
1261 TYCVDPTPCFSCSSLSLSAEDIGCNOTOEAASANTLOIAIEIKIGITRSADPV 1320
1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAVEFPSSGAKSPSKSGAQTPKSPPEHYVQET 1380
1321 SEVPAYSQHPRTKSSRLQSSLSSESARHKAVEFPSSGAKSPSKSGAQTPKSPPEHYVQET 1380
1381 PLMFSCSTSVSLSDSFESRSIASSVQSEPCSGMVSGIISPSDLPSFGQTMPPSRKTPP 1440
1381 PLMFSCSTSVSLSDSFESRSIASSVQSEPCSGMVSGIISPSDLPSFGQTMPPSRKTPP 1440
1441 PPQTQATKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPADTLLHFATESPDGF 1500
1441 PPQTQATKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPADTLLHFATESPDGF 1500
1501 SCSSLSALSLEDPFTQKDVLRIMPVQENDNGNETESEQPKESNENOEKAEKTI DSE 1560
1501 SCSSLSALSLEDPFTQKDVLRIMPVQENDNGNETESEQPKESNENOEKAEKTI DSE 1560
1561 KDLLDSDDDDDIIELEECIIISAMPTKSSRGKKAPOATASKLPPVARKPSQLPVYKLLPS 1620
1561 KDLLDSDDDDDIIELEECIIISAMPTKSSRGKKAPOATASKLPPVARKPSQLPVYKLLPS 1620
1621 QNRLOQKHVSTFPGDDMPRVYCVGPTPIINFSTATSLSDLTIESPPNELAAGEVGGGAQ 1680
1621 QNRLOQKHVSTFPGDDMPRVYCVGPTPIINFSTATSLSDLTIESPPNELAAGEVGGGAQ 1680
1681 SGSEFEKGDITPEGRSTDEAOGKTSSTVITPELDNKAEGDILAEIINSAMPKGSHKP 1740
1681 SGSEFEKGDITPEGRSTDEAOGKTSSTVITPELDNKAEGDILAEIINSAMPKGSHKP 1740
1741 FRVKIMDVOQQAASASSAPNKNQLDGKKKPTSPVKPIQONTEYTRVTRKNADSKNNLN 1800
1741 FRVKIMDVOQQAASASSAPNKNQLDGKKKPTSPVKPIQONTEYTRVTRKNADSKNNLN 1800
1801 AEEVFDNDKSKQNLKNNSKDFNDKLPNNEDVRGSAFDPSPHHYTPTEGTPYCFSRND 1860
1801 AEEVFDNDKSKQNLKNNSKDFNDKLPNNEDVRGSAFDPSPHHYTPTEGTPYCFSRND 1860
1861 SLSSLDFFDDDDVLSREKAEIRKAKENKSEAKVTSTHETLTSNQOSANKTQALAKOPINR 1920

```

```

Db 1861 SLSSLDFFDDVLSREKAEELKAKENKSEAKVTSHTLTSNOQSANKTQAIKQPINR 1920
Qy 1921 GQKPILOKQSTPQSSKOIPDRGAATDEKLNFAJENTVPCFHSNSSLSSISDIDQENN 1980
Db 1921 GQKPILOKQSTPQSSKOIPDRGAATDEKLNFAJENTVPCFHSNSSLSSISDIDQENN 1980
Qy 1981 NKNEPIKETEPDPSOGESKPKQASGVAPKSFHVEDTPVCFGRNSSLSSISDSDDLQ 2040
Db 1981 NKNEPIKETEPDPSOGESKPKQASGVAPKSFHVEDTPVCFGRNSSLSSISDSDDLQ 2040
Qy 2041 ECISAMPKKKPSRLKGNEXHSPRNMGIIGEDITLKDIOIRPDSHGLSPSENPD 2100
Db 2041 ECISAMPKKKPSRLKGNEXHSPRNMGIIGEDITLKDIOIRPDSHGLSPSENPD 2100
Qy 2101 WKAIQEGANSIVSLHQAACCLSQASDSDSILSLKSGISLSPFHLTPDQEKPT 2160
Db 2101 WKAIQEGANSIVSLHQAACCLSQASDSDSILSLKSGISLSPFHLTPDQEKPT 2160
Qy 2161 SNKGPRILKPEKSTLETKKIESKGIKGGKVKYKSLITGKVRNSSEISGQWKQPLQAN 2220
Db 2161 SNKGPRILKPEKSTLETKKIESKGIKGGKVKYKSLITGKVRNSSEISGQWKQPLQAN 2220
Qy 2221 MPISIRGRTHIHPGVNNSSTSPVSKGPPKTPASKSPSEGQTATTSPRGAKPSVKS 2280
Db 2221 MPISIRGRTHIHPGVNNSSTSPVSKGPPKTPASKSPSEGQTATTSPRGAKPSVKS 2280
Qy 2281 ELSPVARQTSQIGSSKAPSRGSRDSTSPRPAQQLSPRIQSPGRNSISPGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGSSKAPSRGSRDSTSPRPAQQLSPRIQSPGRNSISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSPASTKSSGSKMYSYTPGRQMSQNLTKQGLSKXASSIPRSESASKG 2400
Db 2341 KLSQLPRTSPASTKSSGSKMYSYTPGRQMSQNLTKQGLSKXASSIPRSESASKG 2400
Qy 2401 LNQWNGANGANKVELSRMSTKSSGSDSRPVLVQSTFIKEAPSPTRLRKLKLESA 2460
Db 2401 LNQWNGANGANKVELSRMSTKSSGSDSRPVLVQSTFIKEAPSPTRLRKLKLESA 2460
Qy 2461 SPESLSPSPSPASPTRSQAQTPVLSPLPMSLSTHSSVOAGWRKLPNLSPTIYNDG 2520
Db 2461 SPESLSPSPSPASPTRSQAQTPVLSPLPMSLSTHSSVOAGWRKLPNLSPTIYNDG 2520
Qy 2521 RPAKRHDIAKSHSPSRLPINSGTWKRHSKHSSSLPRVSTWRRTGSSSSITLASSBS 2580
Db 2521 RPAKRHDIAKSHSPSRLPINSGTWKRHSKHSSSLPRVSTWRRTGSSSSITLASSBS 2580
Qy 2581 SEXAKSEDEKXVANSISGTKOSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGABS 2640
Db 2581 SEXAKSEDEKXVANSISGTKOSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGABS 2640
Qy 2641 KTLIYQAPAVSKTEDVWRIEDCPINPRSGSPGTGNTPPVSDVSEKANPNIKSDKN 2700
Db 2641 KTLIYQAPAVSKTEDVWRIEDCPINPRSGSPGTGNTPPVSDVSEKANPNIKSDKN 2700
Qy 2701 QAKQNVGSGVPVRTVGLNRLTSFICQDAPDQKGTIKEKQNNPVPVSETNESPIVERT 2760
Db 2701 QAKQNVGSGVPVRTVGLNRLTSFICQDAPDQKGTIKEKQNNPVPVSETNESPIVERT 2760
Qy 2761 PFSSSSSKHSSPSGTVAAVTPFNPNPRSKSADSTGARPQIPTPNNNTKKEDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAAVTPFNPNPRSKSADSTGARPQIPTPNNNTKKEDSKT 2820
Qy 2821 DSTESSGTQSPKHSGLVTSV 2843
Db 2821 DSTESSGTQSPKHSGLVTSV 2843

```

RESULT 12
 US-09-136-605-7
 ; Sequence 7, Application us/09136605A
 ; Patent No. 6140052
 ; GENERAL INFORMATION:
 ; APPLICANT: He, Tong-Chuan

```

; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCP-4, and APC Interact to
; FILE OF INVENTION: Prevent Cancer
; CURRENT APPLICATION NUMBER: US/09/136,605A
; CURRENT FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821,355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003,687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-136-605-7

Query Match 99.7%; Score 14533; DB 3; Length 2973;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2836; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAAASYDQLLKQVEALQWENSILROELEDNSNHLTKLETEASNKVELKQLOQSIIDEAM 60
Db 1 MAAASYDQLLKQVEALQWENSILROELEDNSNHLTKLETEASNKVELKQLOQSIIDEAM 60
Qy 61 ASSQIDILLERLKEINLDSSNFPQVKLRSMKLSRYSYSGREGSVSSRSGECPVPMGSFPR 120
Db 61 ASSQIDILLERLKEINLDSSNFPQVKLRSMKLSRYSYSGREGSVSSRSGECPVPMGSFPR 120
Qy 121 RGFVNGRESTGYLEELKERSLLADLDBEKEKDYIAQLOQLNFKRIDSLPTENFSL 180
Db 121 RGFVNGRESTGYLEELKERSLLADLDBEKEKDYIAQLOQLNFKRIDSLPTENFSL 180
Qy 181 QDTRQLQLEVEARQIRVAMEEQLGTQDMKKAQRRIARIQIEKDIRIQLQLOQAT 240
Db 181 QDTRQLQLEVEARQIRVAMEEQLGTQDMKKAQRRIARIQIEKDIRIQLQLOQAT 240
Qy 241 EAERSSQNKHETGSHDAERONEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 300
Db 241 EAERSSQNKHETGSHDAERONEGQVGEINMATSNGCGSTTRMDHETASVLSSTHSA 300
Qy 301 PRRLTSLHGTVMVYSLMLGTHDKDDMSRTLLAMSSSDSCISMRSQGCCLPLIQLL 360
Db 301 PRRLTSLHGTVMVYSLMLGTHDKDDMSRTLLAMSSSDSCISMRSQGCCLPLIQLL 360
Qy 361 HGNDKDSVLLGNSRSGSEARASAAALHNTIHSQDDPKGRREIRVLHLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNSRSGSEARASAAALHNTIHSQDDPKGRREIRVLHLEQIRAYCETC 420
Qy 421 WEQEAHEPGDQDKNPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAELLQ 480
Db 421 WEQEAHEPGDQDKNPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGQIAELLQ 480
Qy 481 VDCEMYGLTDHYDITLRRYAGMALTNLTGFDVANKATLCSMKGCMRALVAQLKSESED 540
Db 481 VDCEMYGLTDHYDITLRRYAGMALTNLTGFDVANKATLCSMKGCMRALVAQLKSESED 540
Qy 541 QQVIASVLRNLSPADVNSKKTLEBVSVKALMECALEVKESTLKSVALNLSAHCT 600
Db 541 QQVIASVLRNLSPADVNSKKTLEBVSVKALMECALEVKESTLKSVALNLSAHCT 600
Qy 601 ENKADICAVDGAFLVGTTLTYRSQTNTLAIISGGGILNVSLSLIATNEDHRQILRENN 660
Db 601 ENKADICAVDGAFLVGTTLTYRSQTNTLAIISGGGILNVSLSLIATNEDHRQILRENN 660
Qy 661 CLQTLLOHLKSHSLTIIVSNACGTLWNLSARNPKQOEALWDMGAVSMKLNLIHSHKHMAM 720
Db 661 CLQTLLOHLKSHSLTIIVSNACGTLWNLSARNPKQOEALWDMGAVSMKLNLIHSHKHMAM 720
Qy 721 GSAALRNLMANRPKADANIMSPGSSLPVSHVRKQKALEAEIDAQHLSETFDNIDNLS 780

```


721 GSAAALRLMANRPAKYKADANINSGSLPSLHVRRKQKALEAEALDAQHSETFDNIDNLS 780
781 PKASHRSKORHQSLYGDYVFTDNRHDDNSDNFTGNMTVLSPYANTVLPSSSSRG 840
781 PKASHRSKORHQSLYGDYVFTDNRHDDNSDNFTGNMTVLSPYANTVLPSSSSRG 840
841 LDSRSEKORSLERIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEVSAIHTS 900
841 LDSRSEKORSLERIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEVSAIHTS 900
901 QEDRSSGTTLHCVTDERNALRSSAAHSHSNFTKSENSTCSMPYAKLEYKRSS 960
901 QEDRSSGTTLHCVTDERNALRSSAAHSHSNFTKSENSTCSMPYAKLEYKRSS 960
961 NDSLNSVSSNDGYKRGQMKPSIESYSEDDSKFCYQCPADLAHKIHSANHMDNDGE 1020
961 NDSLNSVSSNDGYKRGQMKPSIESYSEDDSKFCYQCPADLAHKIHSANHMDNDGE 1020
1021 LDTINTYLSKYSDQLNSGROSQONERWAPKHIIEDEIKOSQOROSNQSTIYPYTE 1080
1021 LDTINTYLSKYSDQLNSGROSQONERWAPKHIIEDEIKOSQOROSNQSTIYPYTE 1080
1081 STDJXHLKFQHFQOQECVSPYRGRGANGSETNVRGNGHINQVNSQSLCEDDYEDDKP 1140
1081 STDJXHLKFQHFQOQECVSPYRGRGANGSETNVRGNGHINQVNSQSLCEDDYEDDKP 1140
1141 TNYSERVSEEBQHEERPTNYSIKYNEEKHVDPQIDYSLKYATDIPSSQKOSFSFSKS 1200
1141 TNYSERVSEEBQHEERPTNYSIKYNEEKHVDPQIDYSLKYATDIPSSQKOSFSFSKS 1200
1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAOSRSGOPOKAATCKVSSINQETIQ 1260
1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAOSRSGOPOKAATCKVSSINQETIQ 1260
1261 TYCVEDTPICFSRCSLSLSAASDEIGCNQTTQEADSANTLOIAEIKIGITSAEDPV 1320
1261 TYCVEDTPICFSRCSLSLSAASDEIGCNQTTQEADSANTLOIAEIKIGITSAEDPV 1320
1321 SEVFAVSOHPTKSRLOGSSLSSESARHKAVEFPSCAKSPKSGAOTPKSPPHYVOET 1380
1321 SEVFAVSOHPTKSRLOGSSLSSESARHKAVEFPSCAKSPKSGAOTPKSPPHYVOET 1380
1381 PLMFSRCTSVSLSDFSRSSTASSVQSEPCSGMWGIIISPSDLDPSPQWMPPSRSKTPP 1440
1381 PLMFSRCTSVSLSDFSRSSTASSVQSEPCSGMWGIIISPSDLDPSPQWMPPSRSKTPP 1440
1441 PPQTAQKREVPKNKAPTAEKRSQKQAAVNAVQVLPDADTLLHPATSTPDGF 1500
1441 PPQTAQKREVPKNKAPTAEKRSQKQAAVNAVQVLPDADTLLHPATSTPDGF 1500
1501 SCSSLSALSDEPIQDXVELRIMPVQENDNGNETESEOPKESNENOEKAEKTTDSE 1560
1501 SCSSLSALSDEPIQDXVELRIMPVQENDNGNETESEOPKESNENOEKAEKTTDSE 1560
1561 XDLDDSDDDIEILEECIIISAMPTKSRKQKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1561 XDLDDSDDDIEILEECIIISAMPTKSRKQKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1621 QNRLOPQKHVSFTDQDDMPRYVCVEGTPIINSTATSLDLTIESPPNELAAGEVGRGAQ 1680
1621 QNRLOPQKHVSFTDQDDMPRYVCVEGTPIINSTATSLDLTIESPPNELAAGEVGRGAQ 1680
1681 SGEFEKRTDITPEGRSTDEAOGGKTSVITIPELDDNKAEEGDIIAECINSAMPKGSHPK 1740
1681 SGEFEKRTDITPEGRSTDEAOGGKTSVITIPELDDNKAEEGDIIAECINSAMPKGSHPK 1740
1741 FRVKIMQVOQOASASSAPKNQLDGKKKFTSPVKPIPQNTYRTRVRKNADSKNLN 1800
1741 FRVKIMQVOQOASASSAPKNQLDGKKKFTSPVKPIPQNTYRTRVRKNADSKNLN 1800
1801 AERFSDNKNDSKQNLKNSKDFNDKLPNNEDRVGRGSFAFDSPHYHTPIEGTVCFSRND 1860
1801 AERFSDNKNDSKQNLKNSKDFNDKLPNNEDRVGRGSFAFDSPHYHTPIEGTVCFSRND 1860

RESULT 13
US-07-741-940-7
; Sequence 7, Application US/07741940

1861 SLSSLDFFDDDDVLDGREKAEIRKAKENKESAKVTSHTELTNSQOSANKTOAIKOPINR 1920
1861 SLSSLDFFDDDDVLDGREKAEIRKAKENKESAKVTSHTELTNSQOSANKTOAIKOPINR 1920
1921 GPKPILQKQSTFPOSSKDIIDRGAAATDEKLOFAIENTPVCFSHNSLSLSLSDIDQENN 1980
1921 GPKPILQKQSTFPOSSKDIIDRGAAATDEKLOFAIENTPVCFSHNSLSLSLSDIDQENN 1980
1981 NKNEPDKETEPDSDGSEPKQASQYAPKSPHVEDTVPVCFSRNSLSLSLSDIDQENLLQ 2040
1981 NKNEPDKETEPDSDGSEPKQASQYAPKSPHVEDTVPVCFSRNSLSLSLSDIDQENLLQ 2040
2041 ECISSAMPKKKPSRLKGDNEKHSRNMGGIIGEDLTLDLKDIQRPDSHGSPSENED 2100
2041 ECISSAMPKKKPSRLKGDNEKHSRNMGGIIGEDLTLDLKDIQRPDSHGSPSENED 2100
2101 WKAIQEGANSIVSSLHQAAAAACLSQASSSDSDLSLSKSGISLSPHLLTPDQEKPT 2160
2101 WKAIQEGANSIVSSLHQAAAAACLSQASSSDSDLSLSKSGISLSPHLLTPDQEKPT 2160
2161 SNKGPRILKPGKSKTLETTKIESEKGIKGGKVKYKSLITGKVRNSSEISGGMKQPLQAN 2220
2161 SNKGPRILKPGKSKTLETTKIESEKGIKGGKVKYKSLITGKVRNSSEISGGMKQPLQAN 2220
2221 MPSISGRMTMIHPGVNRSSSTSPVSKGPKLTPKSPSPSEGGTATTPSGAKPSVKS 2280
2221 MPSISGRMTMIHPGVNRSSSTSPVSKGPKLTPKSPSPSEGGTATTPSGAKPSVKS 2280
2281 ELSPVARQTSQIGGSSKAPSRGSDSTPSRAQOPLSRPIOSPGRNSISPGRNGISPN 2340
2281 ELSPVARQTSQIGGSSKAPSRGSDSTPSRAQOPLSRPIOSPGRNSISPGRNGISPN 2340
2341 KLSQLPRTSPSTASTKSSGSGKMSYTSRQMSQOQLTKOTGLSKNASSIPRSESASG 2400
2341 KLSQLPRTSPSTASTKSSGSGKMSYTSRQMSQOQLTKOTGLSKNASSIPRSESASG 2400
2401 LNMNNGANGANKVELSRMSSTKSSGESDRSERPVLVQSTFIKEAPSTILRKLKLESA 2460
2401 LNMNNGANGANKVELSRMSSTKSSGESDRSERPVLVQSTFIKEAPSTILRKLKLESA 2460
2461 SFESLSPSRSPASPTRSQACTPVLSPSLPMSLSTSHSSVQAGGWRKLPNLPTIEYNDG 2520
2461 SFESLSPSRSPASPTRSQACTPVLSPSLPMSLSTSHSSVQAGGWRKLPNLPTIEYNDG 2520
2521 RPAKRHDIAHSHSESRLPINRSGTWKREKSHSSLPVSTWRTTSGSSSTLSASSES 2580
2521 RPAKRHDIAHSHSESRLPINRSGTWKREKSHSSLPVSTWRTTSGSSSTLSASSES 2580
2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQVSSGATNGAES 2640
2581 SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQVSSGATNGAES 2640
2641 KTLIIYQVAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVEKANPNIKDSKN 2700
2641 KTLIIYQVAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVEKANPNIKDSKN 2700
2701 QAKQNGVSGVPMRTVGLNRLTSPFOVDAPOKGTIKPGQNNPVVSETNESPIVERT 2760
2701 QAKQNGVSGVPMRTVGLNRLTSPFOVDAPOKGTIKPGQNNPVVSETNESPIVERT 2760
2761 PFSSSSSKHSGSPGTVAAARVTFNYPNPRKSSADSTGARSQIPTPVNNNTKKEDSKT 2820
2761 PFSSSSSKHSGSPGTVAAARVTFNYPNPRKSSADSTGARSQIPTPVNNNTKKEDSKT 2820
2821 DSTESSGTSQPKHSGSYLVTSV 2843
2821 DSTESSGTSQPKHSGSYLVTSV 2843

Patent No. 5352775
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THILVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, McKie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/741,940
 FILING DATE: 19920109
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.035574
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2842 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC

Query Match 99.6%; Score 14515.5; DB 1; Length 2842;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2835; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy	1	MAAASVDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKKEVLKQLQGSIEDEAM	60
Db	1	MAAASVDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNKKEVLKQLQGSIEDEAM	60
Qy	61	ASSGQIDLLERLKLNDSSNPFQVKLRKMSLRSGSREGSVSRSGCSVPVWGSPFR	120
Db	61	ASSGQIDLLERLKLNDSSNPFQVKLRKMSLRSGSREGSVSRSGCSVPVWGSPFR	120
Qy	121	RGVNGSRESTGYLELEKERSLLDLKKEEKDWWYQAQNLTKRIDSLPLTENFSL	180
Db	121	RGVNGSRESTGYLELEKERSLLDLKKEEKDWWYQAQNLTKRIDSLPLTENFSL	179
Qy	181	QTLTRQLEYEYARQIRVAMEEQLGTCDMEKRAQRRIARIQOIEKDIIRIQLQSQAT	240
Db	180	QTMTRQLEYEYARQIRVAMEEQLGTCDMEKRAQRRIARIQOIEKDIIRIQLQSQAT	239
Qy	241	EAERSQNHETGSHDAERQNEQGVGEINMATSNGQGSGTTRMDHETASVLSSSSTHSA	300

Db	240	EAERSQNHETGSHDAERQNEQGVGEINMATSNGQGSGTTRMDHETASVLSSSSTHSA	299
Qy	301	PRRLTSHLGTKEVMVYLLSLMGLTHDKDDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL	360
Db	300	PRRLTSHLGTKEVMVYLLSLMGLTHDKDDMSRTLLAMSSSDSCISMRQSGCLPLLIQLL	359
Qy	361	HGNDKDSVLLGNSRSGKEARASAAALHNIHSDPDKRGRREIRVLHLLLEQIRAYCETC	420
Db	360	HGNDKDSVLLGNSRSGKEARASAAALHNIHSDPDKRGRREIRVLHLLLEQIRAYCETC	419
Qy	421	WEQEAHEPGMDQDNMPAPVEHOICPAVCVLMKLSFDEEHBHNMELGGLQAIALLQ	480
Db	420	WEQEAHEPGMDQDNMPAPVEHOICPAVCVLMKLSFDEEHBHNMELGGLQAIALLQ	479
Qy	481	VDCEMYGLTNDHYSITLRRYAGMALTNLTGFDVANKATLCMKGCMRALVAQLKSESED	540
Db	480	VDCEMYGLTNDHYSITLRRYAGMALTNLTGFDVANKATLCMKGCMRALVAQLKSESED	539
Qy	541	QOVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHC	600
Db	540	QOVIASVRLNLSWRADVNSKTLREVGSVKALMECALEVKKESTLKSLSALNLSAHC	599
Qy	601	ENKADICAVDGAFLVGTTLTYRSQTNLTALIESGGILRNVSLLIATNEDHRQILRENN	660
Db	600	ENKADICAVDGAFLVGTTLTYRSQTNLTALIESGGILRNVSLLIATNEDHRQILRENN	659
Qy	661	CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALWDMGAVSMKNIHSHKHMIAM	720
Db	660	CLQTLLOHLKSHSLTIVSNACGLTNLSARNPKDQBALWDMGAVSMKNIHSHKHMIAM	719
Qy	721	GSAALNLMANRPKYKDANIMSPGSSLSPLSHVRKOKALEAELDAQHLSSETDNDLS	780
Db	720	GSAALNLMANRPKYKDANIMSPGSSLSPLSHVRKOKALEAELDAQHLSSETDNDLS	779
Qy	781	PKASHRSKQRHKQSLGYDYVDFNRRDDNRSDFNTGNMTVLSPYLTNTVLPSSSSSRGS	840
Db	780	PKASHRSKQRHKQSLGYDYVDFNRRDDNRSDFNTGNMTVLSPYLTNTVLPSSSSSRGS	839
Qy	841	LDSSRSKDRSLERERIGIGNYHPATENPGTSKGLQISTTAAQIAKMEVESAHTS	900
Db	840	LDSSRSKDRSLERERIGIGNYHPATENPGTSKGLQISTTAAQIAKMEVESAHTS	899
Qy	901	QEDRSSGTTTLCVTDERNALRRSSAAHTSHNTYNTFTKSENRTCSMPYAKLEYKRSS	960
Db	900	QEDRSSGTTTLCVTDERNALRRSSAAHTSHNTYNTFTKSENRTCSMPYAKLEYKRSS	959
Qy	961	NDSLNSVSSNDGYGKQGMKPSIESYSEDDDESFCYQYPADLAHKIHSANHMDNDGE	1020
Db	960	NDSLNSVSSNDGYGKQGMKPSIESYSEDDDESFCYQYPADLAHKIHSANHMDNDGE	1019
Qy	1021	LDTPINTSLKYSDQLNSGRQSPQNERWARPKHIIEDEIKQSEQRQSRNQSTTYPVTE	1080
Db	1020	LDTPINTSLKYSDQLNSGRQSPQNERWARPKHIIEDEIKQSEQRQSRNQSTTYPVTE	1079
Qy	1081	STDDKHLKFPHFQGCVCVPSYRSRGANGSETNRVSGNHGINSQSLOQEDDYDDKP	1140
Db	1080	STDDKHLKFPHFQGCVCVPSYRSRGANGSETNRVSGNHGINSQSLOQEDDYDDKP	1139
Qy	1141	TNYSERYSEBQHEEBEERTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFQFSKS	1200
Db	1140	TNYSERYSEBQHEEBEERTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFQFSKS	1199
Qy	1201	SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQTIQ	1260
Db	1200	SSGQSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCKVSSINQTIQ	1259
Qy	1261	TYCVEDTPICFRSCSSLSLSAEDDEIGCNOTTCQADSAANTLOIAETKIGITRSADPV	1320
Db	1260	TYCVEDTPICFRSCSSLSLSAEDDEIGCNOTTCQADSAANTLOIAETKIGITRSADPV	1319
Qy	1321	SEVPASVQHPRTKSRSLQSSLSSESARHKAVERFPSSGAKSPSKGAQTPKSPPEHYVQET	1380
Db	1320	SEVPASVQHPRTKSRSLQSSLSSESARHKAVERFPSSGAKSPSKGAQTPKSPPEHYVQET	1379

QY 1381 PLMFRCSTSVSLDSFERSIASSVQSEPCSGMWGIIISPSDLPSDPOQTWPPRSKSTPP 1440
DB 1380 PLMFRCSTSVSLDSFERSIASSVQSEPCSGMWGIIISPSDLPSDPOQTWPPRSKSTPP 1439
QY 1441 PPQTAQTKREVFKNAKTAERKESGPKQAANAAVQVQVLPDADTLLHFAFATSTPDGF 1500
DB 1440 PPQTAQTKREVFKNAKTAERKESGPKQAANAAVQVQVLPDADTLLHFAFATSTPDGF 1499
QY 1501 SCSSLSALSLEDEPIQKDELIMPPVQENDNGNETESEOPKESNENOEKAEKTIIDSE 1560
DB 1500 SCSSLSALSLEDEPIQKDELIMPPVQENDNGNETESEOPKESNENOEKAEKTIIDSE 1559
QY 1561 KOLLDDSDDDDIIEECIIISAMPTKSRRKKPAQATASKLPPPVVARPSOLPVYKLLPS 1620
DB 1560 KOLLDDSDDDDIIEECIIISAMPTKSRRKKPAQATASKLPPPVVARPSOLPVYKLLPS 1619
QY 1621 QNRLOPQKHVSTPDDDDPRVYCVGEGTPIINSTATSLDLETIESPPNELAAGEVGRGAQ 1680
DB 1620 QNRLOPQKHVSTPDDDDPRVYCVGEGTPIINSTATSLDLETIESPPNELAAGEVGRGAQ 1679
QY 1681 SGEFEKRDITPEGRSTDEAOGKTSVVTIPELDDNKAEEGDIILAEICINSAMPKGSHP 1740
DB 1680 SGEFEKRDITPEGRSTDEAOGKTSVVTIPELDDNKAEEGDIILAEICINSAMPKGSHP 1739
QY 1741 FRVKIMDOVQOASASSAPKNQKLDGKKKTSVPKPIQONTYRTVRKNADSKNNLN 1800
DB 1740 FRVKIMDOVQOASASSAPKNQKLDGKKKTSVPKPIQONTYRTVRKNADSKNNLN 1799
QY 1801 AERFVSDNKKQKQKXNKKDFNDKLPNNEDRVGSGPAFDPSPHHYTTPIEGTTPYCFSRND 1860
DB 1800 AERFVSDNKKQKQKXNKKDFNDKLPNNEDRVGSGPAFDPSPHHYTTPIEGTTPYCFSRND 1859
QY 1861 SLSSLDFFDDDDVLDGREKAEKAKENKESAKVTSHTELTSNQQOSANKTOAIKQPNR 1920
DB 1860 SLSSLDFFDDDDVLDGREKAEKAKENKESAKVTSHTELTSNQQOSANKTOAIKQPNR 1919
QY 1921 GQPKTILQKQSTFPOSSKDIIDRGATDEKLNQFAIENTPVCFSNSSLSDIDQENN 1980
DB 1920 GQPKTILQKQSTFPOSSKDIIDRGATDEKLNQFAIENTPVCFSNSSLSDIDQENN 1979
QY 1981 NKENEPIKETEPDQSGEPKQASGYAPKSPHVEDTTPVCFSRNSSLSDIDSEDDLLQ 2040
DB 1980 NKENEPIKETEPDQSGEPKQASGYAPKSPHVEDTTPVCFSRNSSLSDIDSEDDLLQ 2039
QY 2041 ECISAMPKKKXPSBLKGNEXHSPRNMGGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2100
DB 2040 ECISAMPKKKXPSBLKGNEXHSPRNMGGILGEDLTLDLKDIOQPDSEHGLSPDSENF 2099
QY 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSILSLKSGISLGSPPHLPDQEKKEPT 2160
DB 2100 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSILSLKSGISLGSPPHLPDQEKKEPT 2159
QY 2161 SNKGPRILKPGKSTLETKKIESKGIKGGKKVYKSLITGKVRNSEISGQMKOPLQAN 2220
DB 2160 SNKGPRILKPGKSTLETKKIESKGIKGGKKVYKSLITGKVRNSEISGQMKOPLQAN 2219
QY 2221 MFSISGRMTIHTPGVRNSSLSTSPVSKGPKLTPASKSPSEGOTATTSPRGAKPSVKS 2280
DB 2220 MFSISGRMTIHTPGVRNSSLSTSPVSKGPKLTPASKSPSEGOTATTSPRGAKPSVKS 2279
QY 2281 ELSPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGISPPN 2340
DB 2280 ELSPVARQTSQIGGSKAPSRGSRDSTPSRPAQPLSRPIQSPGRNISIPGRNGISPPN 2339
QY 2341 KLSQLPRTSSPTASTKSSGSGMWYTSFGROMSQNLTKQTLGSKNASSIPRSESASK 2400
DB 2340 KLSQLPRTSSPTASTKSSGSGMWYTSFGROMSQNLTKQTLGSKNASSIPRSESASK 2399
QY 2401 LNMOMNNGANKKVELSRMSSTKSSGESDRSERPVLVRQSTFIKEAPSPTLRRKLEESA 2460
DB 2400 LNMOMNNGANKKVELSRMSSTKSSGESDRSERPVLVRQSTFIKEAPSPTLRRKLEESA 2459

QY 2461 SFESLSPSRPASPTRSQACTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
DB 2460 SFESLSPSRPASPTRSQACTPVLSPLPMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2519
QY 2521 RPAKRHDIAASHSESRLPINESCTWKREHSHSSSLPRVSTWRRTGSSSSILSASSES 2580
DB 2520 RPAKRHDIAASHSESRLPINESCTWKREHSHSSSLPRVSTWRRTGSSSSILSASSES 2579
QY 2581 SEKAKSEDEKHVANSISGTSQSKENQVSAKGTWKIKENEFSPNTNSTSQTVSSGATNGAES 2640
DB 2580 SEKAKSEDEKHVANSISGTSQSKENQVSAKGTWKIKENEFSPNTNSTSQTVSSGATNGAES 2639
QY 2641 KTLIYQAPAVSKTDDVWRIEDCPINNPRSGSPGTGNTPPVIDSVSEKANPNIKOSKDN 2700
DB 2640 KTLIYQAPAVSKTDDVWRIEDCPINNPRSGSPGTGNTPPVIDSVSEKANPNIKOSKDN 2699
QY 2701 QAKQVNGGSGVPMRTVGLNRLTSFTQVDAPDQKGTIIRPGQNNPVVSESTNESPIVERT 2760
DB 2700 QAKQVNGGSGVPMRTVGLNRLTSFTQVDAPDQKGTIIRPGQNNPVVSESTNESPIVERT 2759
QY 2761 PFSSSSSSKHSRSGTVAARVTPFNPNPRKSSASDSTGAPRSQIPTPVNNNTKKRDSKT 2820
DB 2760 PFSSSSSSKHSRSGTVAARVTPFNPNPRKSSASDSTGAPRSQIPTPVNNNTKKRDSKT 2819
QY 2821 DSTESSGTQSPKRHSGLVLTSV 2843
DB 2820 DSTESSGTQSPKRHSGLVLTSV 2842

RESULT 14
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids

; TYPE: amino acid		Query Match		99.6%; Score 14515.5; DB 1; Length 2842;	
; STRANDEDNESS: single		Best Local Similarity		99.7%; Pred. No. 0;	
; TOPOLOGY: linear		Matches 2835; Conservative		2; Mismatches 5; Indels 1; Gaps 1;	
; MOLECULE TYPE: protein		1		MAAASYDQLLKQVEALKVENSNTROELDNNSHLTKLETSASNKEVILKOLQSGIDEAM	
; ORIGINAL SOURCE:		1		MAAASYDQLLKQVEALKVENSNTROELDNNSHLTKLETSASNKEVILKOLQSGIDEAM	
; ORGANISM: Homo sapiens		61		ASSGQIDLLERLKEINLDSSNFPQVGLRSKMSLRSYSGREGSVSSRGSGCSPVPMGSPFR	
; IMMEDIATE SOURCE:		61		ASSGQIDLLERLKEINLDSSNFPQVGLRSKMSLRSYSGREGSVSSRGSGCSPVPMGSPFR	
; CLONE: APC		121		RGFVNGRESTGYLEELKERSLLADLDKEEKEKDWYIAQLQNLKRIIDSLPLTENFSL	
; US-08-289-548A-7		121		RGFVNGRESTGYLEELKERSLLADLDKEEKEKDWYIAQLQNLKRIIDSL-LTENFSL	
		181		QTDLTRQLEEARQIRVAMEEQGLTQCDMEKKAORRIARIQOIEKIDILIRIQLLOSQAT	
		180		QTDLTRQLEEARQIRVAMEEQGLTQCDMEKKAORRIARIQOIEKIDILIRIQLLOSQAT	
		241		EARSQNKHETGSHDAERQNEGGVGEINMATSNGQGGTTRMDHETASVLSSTHSA	
		240		EARSQNKHETGSHDAERQNEGGVGEINMATSNGQGGTTRMDHETASVLSSTHSA	
		301		PRRLTSHLGTKEVMYSLILMLGTHDMSRTLLAMSSQDSCI SMRGGCLPLLIQLL	
		300		PRRLTSHLGTKEVMYSLILMLGTHDMSRTLLAMSSQDSCI SMRGGCLPLLIQLL	
		361		HGNDKSVLLGNSRGSKEARASAAALHNIHSQDDKRGRRRIIRVLHLLQIRAYCETC	
		360		HGNDKSVLLGNSRGSKEARASAAALHNIHSQDDKRGRRRIIRVLHLLQIRAYCETC	
		421		WEWQEAHEPGMDQKNPMPAPVEHQICPACVCLMKLSFDEHRHAMELGGIQAIAELIQ	
		420		WEWQEAHEPGMDQKNPMPAPVEHQICPACVCLMKLSFDEHRHAMELGGIQAIAELIQ	
		481		VDCMYGLTNDHYGILTRRYAGMALTNLTGFDVANKATLCSMKGCMFALVAQKSESDL	
		480		VDCMYGLTNDHYGILTRRYAGMALTNLTGFDVANKATLCSMKGCMFALVAQKSESDL	
		541		QOVTASVLRNLWRADYNSKTLREVSQVLMALCALEVKESTLKSLSALWNLSAHCT	
		540		QOVTASVLRNLWRADYNSKTLREVSQVLMALCALEVKESTLKSLSALWNLSAHCT	
		601		ENKADI CAVDGALAFVGLTYRSQNTNLAIIESGGGILRNVSLSLATNEDHRQILRENN	
		600		ENKADI CAVDGALAFVGLTYRSQNTNLAIIESGGGILRNVSLSLATNEDHRQILRENN	
		661		CLQTLQLHLSHSITVSNACGLTWNLSARNPKDQEAALWDMGAVSMLKNLIHSHKHMIAM	
		660		CLQTLQLHLSHSITVSNACGLTWNLSARNPKDQEAALWDMGAVSMLKNLIHSHKHMIAM	
		721		GSAALRNLMANRPAPKADANIMSGSLPSLHVRSKQKALEAELDAQHLSSETFDNIDNLS	
		720		GSAALRNLMANRPAPKADANIMSGSLPSLHVRSKQKALEAELDAQHLSSETFDNIDNLS	
		781		PKASHRSKQRHKSGLYGDYVFDNRHDDNRSDFNTGNMTVLSPLYNTTVLPSSSSSRGS	
		780		PKASHRSKQRHKSGLYGDYVFDNRHDDNRSDFNTGNMTVLSPLYNTTVLPSSSSSRGS	
		841		LDSSRSKDSLERERIGIGLVNHPATENPCTSSKGLQISTTAAOIAKMWEEVSAIHTS	
		840		LDSSRSKDSLERERIGIGLVNHPATENPCTSSKGLQISTTAAOIAKMWEEVSAIHTS	
QY	901	QEDRSSGTTTELHCVTDERNALRRSSAAHTSHNTYNTFTKSENSNRTCSMPYAKLEYKRSS	960		
DB	900	QEDRSSGTTTELHCVTDERNALRRSSAAHTSHNTYNTFTKSENSNRTCSMPYAKLEYKRSS	959		
QY	961	NDLSNSVSSNDGYGKRGQMKPSIESYSDDESKSCSYGQYPADLAHKIHSANHMDNDGE	1020		
DB	960	NDLSNSVSSNDGYGKRGQMKPSIESYSDDESKSCSYGQYPADLAHKIHSANHMDNDGE	1019		
QY	1021	LDTPNYSKYSDEBOLNSGROSPQNERWARPKHIIIEIKQSQROSRNOSTTYPVYTE	1080		
DB	1020	LDTPNYSKYSDEBOLNSGROSPQNERWARPKHIIIEIKQSQROSRNOSTTYPVYTE	1079		
QY	1081	STDDKHLAFQHFQOECVSPYRSGANGSETNRVGNHGINQVNSQLOQEDDYEDDKP	1140		
DB	1080	STDDKHLAFQHFQOECVSPYRSGANGSETNRVGNHGINQVNSQLOQEDDYEDDKP	1139		
QY	1141	TNYSERYSEEEHHEERPTNYSIKYNEEKHVDPIDYSLKYATDIPSSQKQSFSSKS	1200		
DB	1140	TNYSERYSEEEHHEERPTNYSIKYNEEKHVDPIDYSLKYATDIPSSQKQSFSSKS	1199		
QY	1201	SSGQSKTEHSSSENSTPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQSTIQ	1260		
DB	1200	SSGQSKTEHSSSENSTPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQSTIQ	1259		
QY	1261	TYCVEDTPICFRSCLSSLSAEDIGCNQTTQEADSANTLQIAEIKGIGTRSAEDPV	1320		
DB	1260	TYCVEDTPICFRSCLSSLSAEDIGCNQTTQEADSANTLQIAEIKGIGTRSAEDPV	1319		
QY	1321	SEVPAVSQHPPTKASRLQSSLSSESARHKAVERPSSAKSPSKSGAOTPKSPPHYQET	1380		
DB	1320	SEVPAVSQHPPTKASRLQSSLSSESARHKAVERPSSAKSPSKSGAOTPKSPPHYQET	1379		
QY	1381	PLMFSRCTSVSLDSQSFSSRSIASSVQSEPCSMVSGIISPSDLDPSPQTMPPSRSKTTP	1440		
DB	1380	PLMFSRCTSVSLDSQSFSSRSIASSVQSEPCSMVSGIISPSDLDPSPQTMPPSRSKTTP	1439		
QY	1441	PPPTAQTKRVVPKNKAPTAEKRESGPKQAANAQVQVLPDADTLHLHFAESTPDGF	1500		
DB	1440	PPPTAQTKRVVPKNKAPTAEKRESGPKQAANAQVQVLPDADTLHLHFAESTPDGF	1499		
QY	1501	SCSSLSLSALSDDEFFIQKVELRIMPVQENDNGNETESEOPEKESNENQKEAKTTDSE	1560		
DB	1500	SCSSLSLSALSDDEFFIQKVELRIMPVQENDNGNETESEOPEKESNENQKEAKTTDSE	1559		
QY	1561	KDLDSDSDDDIEILEECIIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS	1620		
DB	1560	KDLDSDSDDDIEILEECIIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS	1619		
QY	1621	QNRLOPKHVSFTPCDDMPRVYCVGEGTPIFESTATSLDLTIESPPNELAAGEVGRGAQ	1680		
DB	1620	QNRLOPKHVSFTPCDDMPRVYCVGEGTPIFESTATSLDLTIESPPNELAAGEVGRGAQ	1679		
QY	1681	SGEPEKRTDITEGRSTDEAQQGKTSSVTIPELDDNKAEEGDILAEICINSAMPKSHKP	1740		
DB	1680	SGEPEKRTDITEGRSTDEAQQGKTSSVTIPELDDNKAEEGDILAEICINSAMPKSHKP	1739		
QY	1741	FRVKIMDQVQAASASSAPNKNQIDGKKKPTSPVKPIQNTTEYRTRVRKADSKNIN	1800		
DB	1740	FRVKIMDQVQAASASSAPNKNQIDGKKKPTSPVKPIQNTTEYRTRVRKADSKNIN	1799		
QY	1801	AERVFSNDKSKQNLKXNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTYPYCFSRND	1860		
DB	1800	AERVFSNDKSKQNLKXNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTYPYCFSRND	1859		
QY	1861	SLSLDFDDDDVDLSREKAELURKAKENKESAKVTSHTELTSNOQSANKTQAKAQPINR	1920		
DB	1860	SLSLDFDDDDVDLSREKAELURKAKENKESAKVTSHTELTSNOQSANKTQAKAQPINR	1919		
QY	1921	GQPKPILOKQSTFFQSSKDIIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN	1980		
DB	1920	GQPKPILOKQSTFFQSSKDIIPDRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN	1979		

1981 NKENPIKETEPDPSQGPSPKQASGYAPKSPHVEDTVPVCFSRNSSLSLSDSEDDLLQ 2040
1980 NKENPIKETEPDPSQGPSPKQASGYAPKSPHVEDTVPVCFSRNSSLSLSDSEDDLLQ 2039
2041 ECISAMPKPKKPSRLKGNKHSRNNGGILGDLTLKDIOQPDSEHGLSPDSENF 2100
2040 ECISAMPKPKKPSRLKGNKHSRNNGGILGDLTLKDIOQPDSEHGLSPDSENF 2099
2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSILSKSGISIGSPHLPDQEKPF 2160
2100 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSDSILSKSGISIGSPHLPDQEKPF 2159
2161 SNKGPRILKPKGKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSNISGOMKQPOAN 2220
2160 SNKGPRILKPKGKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSNISGOMKQPOAN 2219
2221 MFSISGRGTMIHIPGVNRSSSSTSPVSKKGPPLKTPASKSPSEGGTATTSRGAKPVS 2280
2220 MFSISGRGTMIHIPGVNRSSSSTSPVSKKGPPLKTPASKSPSEGGTATTSRGAKPVS 2279
2281 ELSPVARTQSIGGSKAPSRSGSDSTPSRPAQOPLSRPTOSPGRNSISPRGNGISPPN 2340
2280 ELSPVARTQSIGGSKAPSRSGSDSTPSRPAQOPLSRPTOSPGRNSISPRGNGISPPN 2339
2341 KLSQLPRTSSPTASTKSGSGKMSYTPSGROMSQNLTKOTGLSKNASSIPRSESASK 2400
2340 KLSQLPRTSSPTASTKSGSGKMSYTPSGROMSQNLTKOTGLSKNASSIPRSESASK 2399
2401 LNMNNGANKKVELSRMSSTKSGSDSDSERPVLVRQSTFIKEAPSPILRRKLEBSA 2460
2400 LNMNNGANKKVELSRMSSTKSGSDSDSERPVLVRQSTFIKEAPSPILRRKLEBSA 2459
2461 SPESLSPSRPASPTRSOAQTPLPSLPDMSLTHSSVQAGGWKLPNLSPIETENDG 2520
2460 SPESLSPSRPASPTRSOAQTPLPSLPDMSLTHSSVQAGGWKLPNLSPIETENDG 2519
2521 RPAKRHDIAHSHSPSLPINKRGTWKREKSHSSLPVSTWRTGSSSIILSASSES 2580
2520 RPAKRHDIAHSHSPSLPINKRGTWKREKSHSSLPVSTWRTGSSSIILSASSES 2579
2581 SEKASEDEKHVNSTSGTKQKNOVSAGTWKIKENEFSTNSTSTVSGATNGAES 2640
2580 SEKASEDEKHVNSTSGTKQKNOVSAGTWKIKENEFSTNSTSTVSGATNGAES 2639
2641 KTLIQMAPAVSKTDEWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKAPNIKDSKDN 2700
2640 KTLIQMAPAVSKTDEWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKAPNIKDSKDN 2699
2701 QAKQNVGSGVPMRTVGLNRLTSTFIQVADPQKGTETKPGQNNPVPVSETNESPIVERT 2760
2700 QAKQNVGSGVPMRTVGLNRLTSTFIQVADPQKGTETKPGQNNPVPVSETNESPIVERT 2759
2761 PFSSSSSSKHSPPSGTVAAARVTPENYNPSRKSSADSTARPQIPTPVNNNTKKRDSKT 2820
2760 PFSSSSSSKHSPPSGTVAAARVTPENYNPSRKSSADSTARPQIPTPVNNNTKKRDSKT 2819
2821 DSTESSGTQSPKRHSGSYLVTSV 2843
2820 DSTESSGTQSPKRHSGSYLVTSV 2842

RESULT 15
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 99.6%; Score 14515.5; DB 1; Length 2842;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2835; Conservative 5; Indels 1; Gaps 1;

Qy 1 MAAASYDQLLKQVEALKWNENLNRLQELDNLSNHLTKLETSANNKEVILKQGSIEDEAM 60
Db 1 MAAASYDQLLKQVEALKWNENLNRLQELDNLSNHLTKLETSANNKEVILKQGSIEDEAM 60
Qy 61 ASSGQIDLLERLKLNDSSNFFGVKLRKWSLRSYSGRSGSVSSRSGSPVPMGSFPR 120
Db 61 ASSGQIDLLERLKLNDSSNFFGVKLRKWSLRSYSGRSGSVSSRSGSPVPMGSFPR 120
Qy 121 RGFVNGSRESTGYLEELKERSULLADLDKEKEKDWYQAQLQNLTKRIDSLPTENFSL 180
Db 121 RGFVNGSRESTGYLEELKERSULLADLDKEKEKDWYQAQLQNLTKRIDSL-LTENFSL 179
Qy 181 QTDLTTRQLEYEARQIRVAMEEQIGTCODMEKRAQRIARIQIQTEKDILIRIQLQSQAT 240
Db 180 QTDLTTRQLEYEARQIRVAMEEQIGTCODMEKRAQRIARIQIQTEKDILIRIQLQSQAT 239
Qy 241 EAERSSQNKHETGSHDAERQNEGQVGEINNATSGNGQGGSTTRMDHETASVLSSTHSA 300
Db 240 EAERSSQNKHETGSHDAERQNEGQVGEINNATSGNGQGGSTTRMDHETASVLSSTHSA 299
Qy 301 PRRLTSHLGTGVENYVLSLMLGTHDKDDMSRTLLAMSSQDSCISMRQSGCPLLIQLL 360
Db 300 PRRLTSHLGTGVENYVLSLMLGTHDKDDMSRTLLAMSSQDSCISMRQSGCPLLIQLL 359

QY 361 HGNDKDSVLLGNSRSGSKEARASAAALHNIIHSOPDDKRGREIRVJLHLEQIRAYCETC 420
Db 360 HGNDKDSVLLGNSRSGSKEARASAAALHNIIHSOPDDKRGREIRVJLHLEQIRAYCETC 419
QY 421 WEMQEAHEPGMDQDNMPAPVEHQICPACVCLMKLSFDEBHRHAMNELGGLQAIJLQ 480
Db 420 WEMQEAHEPGMDQDNMPAPVEHQICPACVCLMKLSFDEBHRHAMNELGGLQAIJLQ 479
QY 481 VDCWYGIJTHDYSITLRRYAGMALTNLTGDDVANKATLCSMKGCMRALVAQIKSSEDL 540
Db 480 VDCWYGIJTHDYSITLRRYAGMALTNLTGDDVANKATLCSMKGCMRALVAQIKSSEDL 539
QY 541 QQVIASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVALNLSAHCT 600
Db 540 QQVIASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVALNLSAHCT 599
QY 601 ENKADICAVDICALFLVGTLYTYSQTNLTALIESGGGILRNVSLLIATNEDHRIQILRENN 660
Db 600 ENKADICAVDICALFLVGTLYTYSQTNLTALIESGGGILRNVSLLIATNEDHRIQILRENN 659
QY 661 CLOTLLOHLKSHSLTIYSNACGLTNLNSARNPKDOBALWDMGAVSMLKNLIHSHKHMIAM 720
Db 660 CLOTLLOHLKSHSLTIYSNACGLTNLNSARNPKDOBALWDMGAVSMLKNLIHSHKHMIAM 719
QY 721 GSAAALNLANRPAKYKANDINIMSPGSLPSLHVRKOKALEAELOHSETFDNIDNLS 780
Db 720 GSAAALNLANRPAKYKANDINIMSPGSLPSLHVRKOKALEAELOHSETFDNIDNLS 779
QY 781 PKASHRSKORHKQSLYGDYFDTNHRHDNDSNFNTGNMTVLSPYNTNTVLPSSSSRGS 840
Db 780 PKASHRSKORHKQSLYGDYFDTNHRHDNDSNFNTGNMTVLSPYNTNTVLPSSSSRGS 839
QY 841 LDDSRSEKDSRLERGERGILGNYPATENPGTSSKRGLOISTTAAQIAKYMESVAIHTS 900
Db 840 LDDSRSEKDSRLERGERGILGNYPATENPGTSSKRGLOISTTAAQIAKYMESVAIHTS 899
QY 901 QEDRSSGSTTELHCVTDERNALRRSSAAHSTHNTYFTKSENSNRCTSMPIAKLEYKRS 960
Db 900 QEDRSSGSTTELHCVTDERNALRRSSAAHSTHNTYFTKSENSNRCTSMPIAKLEYKRS 959
QY 961 NDSLNSVSSNDGYGKQOMKPSIESYSEDDSKFCSYGQYPADLAHKIHSANHMDNDGE 1020
Db 960 NDSLNSVSSNDGYGKQOMKPSIESYSEDDSKFCSYGQYPADLAHKIHSANHMDNDGE 1019
QY 1021 LDTPIYNSLYSDEQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1080
Db 1020 LDTPIYNSLYSDEQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSNQSTTYPVYTE 1079
QY 1081 STDDKHLKFPQHFQOQCVSPYRSRANGSETNRVGSNHNINONVCSLCOEDDYEDDKP 1140
Db 1080 STDDKHLKFPQHFQOQCVSPYRSRANGSETNRVGSNHNINONVCSLCOEDDYEDDKP 1139
QY 1141 TNSYERYSEBEQHEEERPTNYSIKNEEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1200
Db 1140 TNSYERYSEBEQHEEERPTNYSIKNEEKRVDPIDYSLKYATDIPSSQKQSFSSKS 1199
QY 1201 SSGOSSKTEHWSSESSENTSPSSNAKQNLHPSAQSRSQOPKAATCKVSIINQETIQ 1260
Db 1200 SSGOSSKTEHWSSESSENTSPSSNAKQNLHPSAQSRSQOPKAATCKVSIINQETIQ 1259
QY 1261 TYCVEDTPICFRCSLSLSLSAEDIGCNQTTQEADSNANTLQIABIKIGITRISAEDPV 1320
Db 1260 TYCVEDTPICFRCSLSLSLSAEDIGCNQTTQEADSNANTLQIABIKIGITRISAEDPV 1319
QY 1321 SEYPAVNSOHRPTKSSRLQCSLSLSAEDIGCNQTTQEADSNANTLQIABIKIGITRISAEDPV 1380
Db 1320 SEYPAVNSOHRPTKSSRLQCSLSLSAEDIGCNQTTQEADSNANTLQIABIKIGITRISAEDPV 1379
QY 1381 PLMFBSRCTSVSSLDSESRSIASSVQSEPCGVMVSGIISPSDLPDPSGQTMPPSRSKTTP 1440
Db 1380 PLMFBSRCTSVSSLDSESRSIASSVQSEPCGVMVSGIISPSDLPDPSGQTMPPSRSKTTP 1439
QY 1441 PPTQAGTKREVVPKNKAPTAKRESGPKQAANAAVQVLPDADTLHLHFATESTPDGF 1500

Db 1440 PPTQAGTKREVVPKNKAPTAKRESGPKQAANAAVQVLPDADTLHLHFATESTPDGF 1499
QY 1501 SCSSLSLSALSDPEPIQKDVLRIMPPVQENDNGNETESEQPKESNENQEKAEKTIJSE 1560
Db 1500 SCSSLSLSALSDPEPIQKDVLRIMPPVQENDNGNETESEQPKESNENQEKAEKTIJSE 1559
QY 1561 KOLLDDSDDDDIIELEECIIISAMPTKSRKOKKPAQATASKLPPPVARKPSOLPYKLLPS 1620
Db 1560 KOLLDDSDDDDIIELEECIIISAMPTKSRKOKKPAQATASKLPPPVARKPSOLPYKLLPS 1619
QY 1621 QNRLQPKQHVFTPCDDMPRVYCVGTPINFSTATSLSDLTIESPPNELAAGEGVRGAQ 1680
Db 1620 QNRLQPKQHVFTPCDDMPRVYCVGTPINFSTATSLSDLTIESPPNELAAGEGVRGAQ 1679
QY 1681 SGEPFKRDTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKGSHPK 1740
Db 1680 SGEPFKRDTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINSAMPKGSHPK 1739
QY 1741 FRVKIMDOVOQOASASSAPNKNOLDGKKKXPTSPVKPIPONTYRTRVRKNADSKNIN 1800
Db 1740 FRVKIMDOVOQOASASSAPNKNOLDGKKKXPTSPVKPIPONTYRTRVRKNADSKNIN 1799
QY 1801 AERVSFNKDSKKQNLKNNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTPYCFSRND 1860
Db 1800 AERVSFNKDSKKQNLKNNSKDFNDKLPNNEDRVGSAFDSPHHYTPIEGTPYCFSRND 1859
QY 1861 SLSLDPDDDDVILSRKAEALRKAKENKESPAKVTSHLTSTNOOSANKTOQAIKAPINR 1920
Db 1860 SLSLDPDDDDVILSRKAEALRKAKENKESPAKVTSHLTSTNOOSANKTOQAIKAPINR 1919
QY 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFPAIENTPVCFSHNSLSLSLSDIDQNN 1980
Db 1920 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFPAIENTPVCFSHNSLSLSLSDIDQNN 1979
QY 1981 NKNEPIKETEPFPPSQEPSPQASGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDQNN 2040
Db 1980 NKNEPIKETEPFPPSQEPSPQASGYAPKSFHVEDTPVCFSRNSSLSLSLSDIDQNN 2039
QY 2041 ECISAMPKPKKSRLLKGDNEKHSRNGGILGSDTLTLDKDIQRPDSEHGLSPDSNFD 2100
Db 2040 ECISAMPKPKKSRLLKGDNEKHSRNGGILGSDTLTLDKDIQRPDSEHGLSPDSNFD 2099
QY 2101 WKALQEGANSTVSSLHOAAAAACLSRQASDSDSILSKGILSGSPFHLTPDOEEKPFT 2160
Db 2100 WKALQEGANSTVSSLHOAAAAACLSRQASDSDSILSKGILSGSPFHLTPDOEEKPFT 2159
QY 2161 SNKGPRILKPKCKSTLETKIETESKGIKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2220
Db 2160 SNKGPRILKPKCKSTLETKIETESKGIKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2219
QY 2221 MPSISRGRTMHIHFGVNRNSSSTSPVSKGPPPLKTPASKSPSEGTATTPRGAAPSVKS 2280
Db 2220 MPSISRGRTMHIHFGVNRNSSSTSPVSKGPPPLKTPASKSPSEGTATTPRGAAPSVKS 2279
QY 2281 ELSPVAKQTSQIGGSSKAPRSRSGRSDTSPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2340
Db 2280 ELSPVAKQTSQIGGSSKAPRSRSGRSDTSPRPAQPLSRPIQSPGRNSISPGRNGISPPN 2339
QY 2341 KLSQLPRTSSPSTASTKSSGSKMSYTPSPQMSQNLTKQTLGSKNASSIPRSESASKG 2400
Db 2340 KLSQLPRTSSPSTASTKSSGSKMSYTPSPQMSQNLTKQTLGSKNASSIPRSESASKG 2399
QY 2401 LNQWNGNANGKVKELSRVMSSTKSSGESDRSEPVLRQSTFIKEAPSTLKRKEESA 2460
Db 2400 LNQWNGNANGKVKELSRVMSSTKSSGESDRSEPVLRQSTFIKEAPSTLKRKEESA 2459
QY 2461 SFESLSFSSRSPASPTRQAQTPVLSPSLPMSLSTSHSVQAGGWRKLPPLNLSPTIENDG 2520
Db 2460 SFESLSFSSRSPASPTRQAQTPVLSPSLPMSLSTSHSVQAGGWRKLPPLNLSPTIENDG 2519
QY 2521 RPAKHDIARSHSSRPLPINSRGTWKREHSKSSSLPRVSTWRPTGSSSSTILSASSES 2580

Db	2520	RPAKRHDIARSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES	2579
Qy	2581	SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES	2640
Db	2580	SEKAKSEDEKHVNSISGTQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES	2639
Qy	2641	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN	2700
Db	2640	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSKDN	2699
Qy	2701	QAKQNVGNGSVPMETVGLNLTSTFIOVDAPDOKGTEIKGQNNPVPVSETNESPIVERT	2760
Db	2700	QAKQNVGNGSVPMETVGLNLTSTFIOVDAPDOKGTEIKGQNNPVPVSETNESPIVERT	2759
Qy	2761	PFSSSSSKHSSPSGTVAARVTPFNYPNPSPKSSADSTARSQIPTPVNNNTKKRDSKT	2820
Db	2760	PFSSSSSKHSSPSGTVAARVTPFNYPNPSPKSSADSTARSQIPTPVNNNTKKRDSKT	2819
Qy	2821	DSTESSGTQSPKRHSGSYLVTSV	2843
Db	2820	DSTESSGTQSPKRHSGSYLVTSV	2842

Search completed: August 25, 2004, 17:25:42
 Job time : 53.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:06:26 ; Search time 59 Seconds
(without alignments)
4635.127 Million cell updates/sec

Title: US-09-442-489F-7
Perfect score: 14566
Sequence: 1 MAASVDQLLKQVEALKMEN.....ESSGTQSPKHSGLVTSV 2843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR.78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14566	100.0	2843	1 RBHUAP	adenomatous polyposis coli protein - human
2	13178	90.5	2845	2 T30336	adenomatous polyposis coli protein - human
3	3512.5	24.1	2274	2 T30258	adenomatous polyposis coli protein - human
4	1780.5	12.2	2415	2 T13825	adenomatous polyposis coli protein - human
5	649	4.5	5327	2 T13584	adenomatous polyposis coli protein - human
6	517	3.5	2271	2 P30073	adenomatous polyposis coli protein - human
7	513	3.5	2232	2 T34434	adenomatous polyposis coli protein - human
8	511.5	3.5	1186	2 T23327	adenomatous polyposis coli protein - human
9	509.5	3.3	1188	2 T23330	adenomatous polyposis coli protein - human
10	476	3.3	2774	2 A43359	adenomatous polyposis coli protein - human
11	454	3.1	3507	2 T34513	adenomatous polyposis coli protein - human
12	453	3.1	4776	2 E95206	adenomatous polyposis coli protein - human
13	426.5	2.9	1791	2 T02345	adenomatous polyposis coli protein - human
14	420	2.9	2938	2 T30249	adenomatous polyposis coli protein - human
15	417.5	2.9	2722	2 T20532	adenomatous polyposis coli protein - human
16	415.5	2.9	2738	2 E88320	adenomatous polyposis coli protein - human
17	409	2.8	3942	2 T42730	adenomatous polyposis coli protein - human
18	405.5	2.8	3488	2 T34418	adenomatous polyposis coli protein - human
19	405	2.8	2187	2 T30826	adenomatous polyposis coli protein - human
20	404	2.8	2688	2 I49477	adenomatous polyposis coli protein - human
21	400	2.7	2526	2 T20531	adenomatous polyposis coli protein - human
22	399	2.7	3924	2 S37431	adenomatous polyposis coli protein - human
23	395.5	2.7	3187	2 T20537	adenomatous polyposis coli protein - human
24	395	2.7	5170	2 T20538	adenomatous polyposis coli protein - human
25	394.5	2.7	3147	2 T18674	adenomatous polyposis coli protein - human
26	394	2.7	6713	2 B99921	adenomatous polyposis coli protein - human
27	393.5	2.7	5105	2 T32650	adenomatous polyposis coli protein - human
28	391.5	2.7	3562	2 A47171	adenomatous polyposis coli protein - human
29	379.5	2.6	2248	2 A35938	adenomatous polyposis coli protein - human

30	378.5	2.6	2253	2 T30336	nuclear/mitotic ap
31	377	2.6	1459	2 T32271	hypothetical prote
32	376.5	2.6	2176	2 T18006	toucan gene protei
33	376	2.6	971	2 T19431	hypothetical prote
34	375.5	2.6	2364	2 A56577	microtubule-associ
35	374.5	2.6	4377	2 A55575	ankyrin 3, long sp
36	374	2.6	3256	2 A48666	cell proliferation
37	372	2.6	2649	2 T51023	hypothetical prote
38	371.5	2.6	1560	2 T02885	peroxisome prolife
39	370.5	2.5	2346	2 T18829	nuclear receptor c
40	370	2.5	2453	2 S60254	Basoon protein -
41	369	2.5	3938	2 T42761	breast cancer tumo
42	367	2.5	3328	2 T30835	trithorax homolog
43	367	2.5	3968	2 A44265	hypothetical prote
44	365	2.5	2218	2 B84683	nucleolar phosphop
45	364.5	2.5	990	2 T51618	

ALIGNMENTS

RESULT 1

RBHUAP
adenomatous polyposis coli protein - human
N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #text_change 21-Jul-2000
C;Accession: A37261, B39658, A49319; I54271
R;Kinler, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith, Chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.
Science 253, 661-665, 1991
A;Title: Identification of FAP locus genes from chromosome 5q21.
A;Reference number: A37261; MUID:91335210; PMID:1651562
A;Accession: A37261
A;Molecule type: mRNA
A;Residues: 1-2843 <KIN>
A;Cross-references: GB:M74088; NID:9182396; PIDN:AAA03586.1; PID:9182397
R;Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grode, arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Leif Cell 66, 601-613, 1991
A;Title: Identification of deletion mutations and three new genes at the familial polyposis A;Reference number: A39658; MUID:91330307; PMID:1678319
A;Accession: B39658
A;Molecule type: DNA
A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P', A;Cross-references: GB:M73548; NID:9190163; PIDN:AAA60354.1; PID:9190164
R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelstein Cancer Res. 52, 643-645, 1992
A;Title: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a c A;Reference number: A44928; MUID:92119623; PMID:1310068
A;Accession: A44928
A;Molecule type: DNA
A;Residues: 1506-1525 <MIK>
A;Cross-references: NID:9243541; PIDN:AAB21145.1; PID:9243542
A;Note: sequence extracted from NCBI backbone (NCBI:78214, NCBIP:78218)
R;Spirio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelbert Cell 75, 951-957, 1993
A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.
A;Reference number: A49319; MUID:94073973; PMID:8252630
A;Accession: A49319
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 'G', 143-171, 'P', 173-179 <SPI>
A;Cross-references: GB:S67787; NID:9461061; PIDN:AAD13997.1; PID:94261697
R;Lambert, S.; Ballhausen, W.G.
Hum. Genet. 90, 650-652, 1993
A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly A;Reference number: I54271; MUID:93186137; PMID:8383094
A;Accession: I54271
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-4 <LAW>
A;Cross-references: GB:S56365; NID:9266243; PIDN:AAD14918.1; PID:94262770

C;Genetics:

A;Gene: GDB:APC
 A;Cross-references: GDB:119682; OMIM:175100
 A;Map position: 5q21-5q22
 A;Note: mutations of this gene can result in familial adenomatous polyposis or sporadic
 C;Superfamily: adenomatous polyposis coli protein
 C;Keywords: cancer; familial adenomatous polyposis; tumor suppressor
 F;1-730/Domain: leucine-rich <NTD>
 F;7-72/Region: coil #status predicted
 F;185-227/Region: coil #status predicted
 F;731-2832/Domain: serine-rich <CTD>
 F;1131-1156/Region: acidic
 F;1558-1577/Region: acidic
 F;1866-1893/Region: highly charged

Query Match 100.0%; Score 14566; DB 1; Length 2843;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAAASVDOLLKQVEALKVENSNIROLEDNSNHLTKYTEASNKKEVLKOLQSGSIEDEAM	60	781	PKASHRSKQRHKQSLYGDYVDFDTNRHDDNRSNDPNFTGNMTVLSPYLNTVLPSSSSRG	840
Db	1	MAAASVDOLLKQVEALKVENSNIROLEDNSNHLTKYTEASNKKEVLKOLQSGSIEDEAM	60	841	LDSSRSKDRSLERERIGLGNVHPATENPCTSSKRGQLISTTAAQIAKYMEEVSAIHTS	900
Qy	61	ASSGQIDLLERLKEALNDSSNFFGVKLRKMSLRYSYSGREGSVSRGSCSPVPMGSPFR	120	841	LDSSRSKDRSLERERIGLGNVHPATENPCTSSKRGQLISTTAAQIAKYMEEVSAIHTS	900
Db	61	ASSGQIDLLERLKEALNDSSNFFGVKLRKMSLRYSYSGREGSVSRGSCSPVPMGSPFR	120	901	QEDRSSGTTTTLCHVCTDERNALRRSSAAHNTSNNTYNTFKSNSNRTCSMPYAKLEYKRSS	960
Qy	121	RGFVNSRESTGYLEBEKERSILLADLKEEKEKDWYIAQONLTKRIDSLPTENFSL	180	901	QEDRSSGTTTTLCHVCTDERNALRRSSAAHNTSNNTYNTFKSNSNRTCSMPYAKLEYKRSS	960
Db	121	RGFVNSRESTGYLEBEKERSILLADLKEEKEKDWYIAQONLTKRIDSLPTENFSL	180	961	NDLSNSVSSSDGYKRGOMKPSIESYSEDDDESKFCSYGOYPADLAHKIHSANHMDNDGE	1020
Qy	181	QDTMTROLEYEARQIRVAMEEOLGTQDMKEKRAQRIARIOIEKDILIRIQLQSOAT	240	961	NDLSNSVSSSDGYKRGOMKPSIESYSEDDDESKFCSYGOYPADLAHKIHSANHMDNDGE	1020
Db	181	QDTMTROLEYEARQIRVAMEEOLGTQDMKEKRAQRIARIOIEKDILIRIQLQSOAT	240	1021	LDTPIYSLKYSDQLNSGRQSPQSNRERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE	1080
Qy	241	EABRSSONKHETGSHDAERQNEGGVGEINMATSGNGQSTTRMDHETASVLSSTHSA	300	1021	LDTPIYSLKYSDQLNSGRQSPQSNRERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE	1080
Db	241	EABRSSONKHETGSHDAERQNEGGVGEINMATSGNGQSTTRMDHETASVLSSTHSA	300	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGSNHNQINQVOSLCEQDDYEDDKP	1140
Qy	301	PRLTSHLGTPKVMYSLLSMLGTHDKDMSRTLLAMSSQSCISMRQSGCLPLLIQLL	360	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGSNHNQINQVOSLCEQDDYEDDKP	1140
Db	301	PRLTSHLGTPKVMYSLLSMLGTHDKDMSRTLLAMSSQSCISMRQSGCLPLLIQLL	360	1141	TNYSERYSSEEEHHEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSPKS	1200
Qy	361	HGNDKDSVLLGNRSGKEARASAAHNIHSPQDDKGRREIRVHLLEQIRAYCETC	420	1141	TNYSERYSSEEEHHEERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKQSFSPKS	1200
Db	361	HGNDKDSVLLGNRSGKEARASAAHNIHSPQDDKGRREIRVHLLEQIRAYCETC	420	1201	SSGQSSKTEHMSSENSTSTPSSNAKQNLHPSSAQSRGQPKAATCKVSSINQETIQ	1260
Qy	421	WEMQEAHEPGMDQKNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQAELLLQ	480	1201	SSGQSSKTEHMSSENSTSTPSSNAKQNLHPSSAQSRGQPKAATCKVSSINQETIQ	1260
Db	421	WEMQEAHEPGMDQKNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQAELLLQ	480	1261	TYCVEDTPICFSRCSLSSLSAEDEIGCQNTTQEADSANTLOIAETKEIGTRSAEDPV	1320
Qy	481	VCCEMYGLTNDHYSITLRYAGMALTNLTFDGVANKATLCMKGWRALVAQLKSESED	540	1261	TYCVEDTPICFSRCSLSSLSAEDEIGCQNTTQEADSANTLOIAETKEIGTRSAEDPV	1320
Db	481	VCCEMYGLTNDHYSITLRYAGMALTNLTFDGVANKATLCMKGWRALVAQLKSESED	540	1321	SEVPAVQCHPRTKSRRLQSSLSSESARHKAVERFSSGAKSPKSGAQTPKSPPEHYVQET	1380
Qy	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALVYKKESTLKSVLASALWNSAHCT	600	1321	SEVPAVQCHPRTKSRRLQSSLSSESARHKAVERFSSGAKSPKSGAQTPKSPPEHYVQET	1380
Db	541	QOVIASVLRNLWRADVNSKTLREVGSVKALMECALVYKKESTLKSVLASALWNSAHCT	600	1381	PLMFSRCTSVSSLDSPSRSIASSVQSPCGMVSGIISPSDLPDPGQTPMPSRSKTPP	1440
Qy	601	ENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILRNYSLSIATNEDHRLQREN	660	1381	PLMFSRCTSVSSLDSPSRSIASSVQSPCGMVSGIISPSDLPDPGQTPMPSRSKTPP	1440
Db	601	ENKADICAVDGAFLVGLTYRSQNTLAIIESGGGILRNYSLSIATNEDHRLQREN	660	1441	PPPTAQTAKREVPKNAKTAKEKESGPKQAAVNAAVQVRQVLPDADTLHLHFAETSTPDGF	1500
Qy	661	CLQTLQHLKSHSLTIVSNACGTLWNLARNPKDQBALWDMGAVSVLKNLHSHKHMIA	720	1441	PPPTAQTAKREVPKNAKTAKEKESGPKQAAVNAAVQVRQVLPDADTLHLHFAETSTPDGF	1500
Db	661	CLQTLQHLKSHSLTIVSNACGTLWNLARNPKDQBALWDMGAVSVLKNLHSHKHMIA	720	1501	SCSSLSALSALDEPFIQKDVLEIRIMPPVQNDNGNETESEQPKESNENQKEAEKTIIDSE	1560
Qy	721	GSAALARNLMANRPAYKDKANTMSPGSSILPSLHVKKQKALELDAQHLSETFDNIDLS	780	1501	SCSSLSALSALDEPFIQKDVLEIRIMPPVQNDNGNETESEQPKESNENQKEAEKTIIDSE	1560
Db	721	GSAALARNLMANRPAYKDKANTMSPGSSILPSLHVKKQKALELDAQHLSETFDNIDLS	780	1561	KDLDDDDDDDDIIELEECIIISAMPTKSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS	1620
Qy	781	PKASHRSKQRHKQSLYGDYVDFDTNRHDDNRSNDPNFTGNMTVLSPYLNTVLPSSSSRG	840	1561	KDLDDDDDDDDIIELEECIIISAMPTKSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS	1620

1921 QGPKPILOKQSTFPQSSKIDPDRGAATDEKIQNPAIENTPVCFSHNSLSLSDIDQENN 1980
 Db 1921 QGPKPILOKQSTFPQSSKIDPDRGAATDEKIQNPAIENTPVCFSHNSLSLSDIDQENN 1980
 Qy 1981 NKENPEIKETEPDPSQCEPKPQASGVAPKSFHVEDTPVCFSRNSSLSLSDIDEDLLQ 2040
 Db 1981 NKENPEIKETEPDPSQCEPKPQASGVAPKSFHVEDTPVCFSRNSSLSLSDIDEDLLQ 2040
 Qy 2041 ECISAMPKPKKPKRLKGDNEKHSFPRNMGITLGBDLTLDLKIQRPDSEHGLSPDSENF 2100
 Db 2041 ECISAMPKPKKPKRLKGDNEKHSFPRNMGITLGBDLTLDLKIQRPDSEHGLSPDSENF 2100
 Qy 2101 WKAIQEGANSTVSLHAAAAACLSROASDSDSLKSGISLGSPFHLTPQEEKPFT 2160
 Db 2101 WKAIQEGANSTVSLHAAAAACLSROASDSDSLKSGISLGSPFHLTPQEEKPFT 2160
 Qy 2161 SNKGPRILKPEKSTLTETKTESKGIKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2220
 Db 2161 SNKGPRILKPEKSTLTETKTESKGIKGGKVKYKSLITGKVRNSSEISGQMKQPLQAN 2220
 Qy 2221 MPSISRGRTMIHIPGVNNSGSSSTSPVSKKGPPLKTPASKSPSEGOATTTPRGAKPSVK 2280
 Db 2221 MPSISRGRTMIHIPGVNNSGSSSTSPVSKKGPPLKTPASKSPSEGOATTTPRGAKPSVK 2280
 Qy 2281 ELSVARTSOTIGSSKAPSGSRDSTPPRPAOQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Db 2281 ELSVARTSOTIGSSKAPSGSRDSTPPRPAOQPLSRPIQSPGRNSISPGRNGISPPN 2340
 Qy 2341 KLSQLPRTSSPTASTKSSGSKNYSYTPGRQMSQQNLTKQTGLSKNASSIPRSESASK 2400
 Db 2341 KLSQLPRTSSPTASTKSSGSKNYSYTPGRQMSQQNLTKQTGLSKNASSIPRSESASK 2400
 Qy 2401 LNQMNGANGANKVELSRMSSTKSGESDRSERPVLVROSTFIKEAPSTLRRKLEESA 2460
 Db 2401 LNQMNGANGANKVELSRMSSTKSGESDRSERPVLVROSTFIKEAPSTLRRKLEESA 2460
 Qy 2461 SFESLSPSSRPASTRSOAQTPVLSPLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
 Db 2461 SFESLSPSSRPASTRSOAQTPVLSPLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG 2520
 Qy 2521 RPAKRHDIAHSHSPSRPLNRSRGTWKRHSKHSLSLPRVSTWRRTGSSSSLSASSES 2580
 Db 2521 RPAKRHDIAHSHSPSRPLNRSRGTWKRHSKHSLSLPRVSTWRRTGSSSSLSASSES 2580
 Qy 2581 SEKAKEDEKHNSTSGTKQSKENQVSAKGTWRKI KENERFPTNSTQTQVSSGATNGAES 2640
 Db 2581 SEKAKEDEKHNSTSGTKQSKENQVSAKGTWRKI KENERFPTNSTQTQVSSGATNGAES 2640
 Qy 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKDN 2700
 Db 2641 KTLIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKDN 2700
 Qy 2701 QAKQVNGSVPMRTVGLNLSNFIQYDAPQKGTETKPKQNNPVPVSETNNESSIVERT 2760
 Db 2701 QAKQVNGSVPMRTVGLNLSNFIQYDAPQKGTETKPKQNNPVPVSETNNESSIVERT 2760
 Qy 2761 PFGSSSSKHSPPSGTVAARTVPFNPNPSRKSSADSTSAEPSQITPPVNNNTKKRDSKT 2820
 Db 2761 PFGSSSSKHSPPSGTVAARTVPFNPNPSRKSSADSTSAEPSQITPPVNNNTKKRDSKT 2820
 Qy 2821 DSTESSGTQSPKRHSGSYLVTSV 2843
 Db 2821 DSTESSGTQSPKRHSGSYLVTSV 2843

RESULT 2
 149505
 adenomatous polyposis coli protein - mouse
 N:Alternate names: APC
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C:Accession: 149505

R;Su, L.
 Science 256, 668-670, 1992
 A>Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the
 A:Reference number: 149505; MUID:92263101; PMID:1350108
 A:Accession: 149505
 A>Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-2845 <RES>
 A:Cross-references: GB:M88127; NID:gl91991; PIDN:AA859632.1; PID:gl91992
 C:Superfamily: adenomatous polyposis coli protein

Query Match 90.5%; Score 13178; DB 2; Length 2845;
 Best Local Similarity 90.2%; Pred. No. 0;
 Matches 2572; Conservative 111; Mismatches 155; Indels 12; Gaps 10;
 Qy 1 MAASVDQLLKQVEALKMENSNLROELEDNSNHLTKLETEASNMEVLKQLGSGIDEAM 60
 Db 1 MAASVDQLLKQVEALKMENSNLROELEDNSNHLTKLETEASNMEVLKQLGSGIDEAM 60
 Qy 61 ASSGQIDLLERLKEINLSDSNFPVKLRKMSLRSGVSGREGSVSRSGECSPVPMGSPFR 120
 Db 61 -TSGQIDLLERLKEINLSDSNFPVKLRKMSLRSGVSGREGSVSRSGECSPVPMGSPFR 118
 Qy 121 RGFVNGSRSTGYLBELEKERSLLADLDKEKEKDWYVAQ-LNLTKRIDSPLTENFSL 180
 Db 119 RGFVNGSRSTGYLBELEKERSLLADLDKEKEKDWYVAQ-LNLTKRIDSPLTENFSL 178
 Qy 181 QDTMTROLEYEARQIRVAMEEQLGTCQDMKEAQRRIARIQOIEKDIILIRIQLLOSQAT 240
 Db 179 QDTMTROLEYEARQIRVAMEEQLGTCQDMKEAQRRIARIQOIEKDIILIRIQLLOSQAA 238
 Qy 241 EAERSQNKHETGSHDAERQNEGGVGEINMATSGNGQSTTMDHETASVLSSSSTHSA 300
 Db 239 EAERSQSRHDAASHEAGQHEGHGVAESNTAASSSGQSPATRVDHETASVLSSSSTHSA 298
 Qy 301 PRRLTSHLGTVMYVYSLLSMLGTHDKDMRSRTLLAMSSQSDSCISMROSGCLPLLIQLL 360
 Db 299 PRRLTSHLGTVMYVYSLLSMLGTHDKDMRSRTLLAMSSQSDSCISMROSGCLPLLIQLL 358
 Qy 361 HGNKDVSLLGNSRSGKEARASAAALHNIHSQDDPKRGRREIRVLHLEQIRACETC 420
 Db 359 HGNKDVSLLGNSRSGKEARASAAALHNIHSQDDPKRGRREIRVLHLEQIRACETC 418
 Qy 421 WEQEAHEPMDQDNMPAPVVEHQICPAVCVLMKLSFDEEHRHANNELGGQAIALLQ 480
 Db 419 WEQEAHEPMDQDNMPAPVVEHQICPAVCVLMKLSFDEEHRHANNELGGQAIALLQ 478
 Qy 481 VDCMYGLTNDHYITLRYAGMALTNLTFGDVANKATLCMKGMCRALVAQIKSEEDL 540
 Db 479 VDCMYGLTNDHYITLRYAGMALTNLTFGDVANKATLCMKGMCRALVAQIKSEEDL 538
 Qy 541 QQVIASVLRNLSWRADVNSKTLREYGVSKALMECALEVKESTLKSVALMNLSAHCT 600
 Db 539 QQVIASVLRNLSWRADVNSKTLREYGVSKALMECALEVKESTLKSVALMNLSAHCT 598
 Qy 601 ENKADICAVDGAALAFVLGTLTVRSQNTLAIIESGGIILRNVSLLIATNEDHQIILRENN 660
 Db 599 ENKADICAVDGAALAFVLGTLTVRSQNTLAIIESGGIILRNVSLLIATNEDHQIILRENN 658
 Qy 661 CLQTLLOHLKSHSLTVSNACGTLNLSARNPKQDALWDMGAVSMLKNLIHSHKHMIAM 720
 Db 659 CLQTLLOHLKSHSLTVSNACGTLNLSARNPKQDALWDMGAVSMLKNLIHSHKHMIAM 718
 Qy 721 GSAALARNLMANRPAPKYKQANIMSPGSSLSFSLHVRKQKALEAELDAQHLSFDPNDLNS 780
 Db 719 GSAALARNLMANRPAPKYKQANIMSPGSSLSFSLHVRKQKALEAELDAQHLSFDPNDLNS 778
 Qy 781 PKASHRSKQKHQSLYGDYVVDNRRDDNRSDNFNTGNNTVLSPLYNTTTLVLPSSSSSRGS 840
 Db 779 PKASHRSKQKHQSLYGDYVVDNRRDDNRSDNFNTGNNTVLSPLYNTTTLVLPSSSSSRGS 838
 Qy 841 LDSRSSEKRSLERRERGIGLGNTHYPATENPGTSSKRGLOISTTAAQIAKYMEVSVSIHTS 900

Db 839 LDSRSRSEKDRSLRERIGIGLSAYHPTTENAGTSSKRGQLQITTTAAQIAKVMREVSALHTS 898
 Qy 901 QDRSSGTTTELHCVTDERNALRRSSAAHSTHNTYNTKSENRTCSMPYAKLEYKRSS 960
 Db 899 QDRSSASTTEFFCVAADRSAARSSASHSTHNTYNTKSENRTCSMPYAKLEYKRSS 958
 Qy 961 NDSLSNVSSSDGKGKGMKPSYESYSEDDSKFCSYGYVPADLAHKIHSANHMDNDGE 1020
 Db 959 NDSLSNVSSSDGKGKGMKPSYESYSEDDSKFCSYGYVPADLAHKIHSANHMDNDGE 1018
 Qy 1021 LDTPIYSLKYDEQLNSGRQSPQNERWAPKHIIIEIKOSEQRQSNQSTTPYVYTE 1080
 Db 1019 LDTPIYSLKYDEQLNSGRQSPQNERWAPKHIIIEIKOSEQRQSNQSTTPYVYTE 1078
 Qy 1081 STDHKLKFPQHFQOQECVSPYRSRGANGSETNVCNSHCINQVNSQSCQEDDVEDDKP 1140
 Db 1079 NTDDKHLKFPQHFQOQECVSPYRSRGANGSETNVCNSHCINQVNSQSCQEDDVEDDKP 1138
 Qy 1141 TNSERYSEBEQH-EBEERTNTSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSPFSK 1199
 Db 1139 TNSERYSEBEQH-EBEERTNTSIKYNEEKHVDQPIDYSLKYATDIPSSQKQSPFSK 1198
 Qy 1200 SSGQSKTHMSSSENTSTPSNAKRONLHPSSAQSRSGOPOKAAATCKVSSINQETI 1259
 Db 1199 NSSAQSTKPEHLSPSENTAVPPSNAKRONLHPSSAQ-RNGQTKGTTCVKPSPINQETI 1257
 Qy 1260 QTCVDETPICFRKCSLSLSAEDEIGCNOTTOEADSAANTLOIAIEIKIKIGTSAEDP 1319
 Db 1258 QTCVDETPICFRKCSLSLSAEDEIGCNOTTOEADSAANTLOIAIEIKIGTSAEDP 1317
 Qy 1320 VSEVPVAVSQHPTKSRLOQSSLSLSARH-KAVEFPSSGAKSPSKSGAQTPKSPPEHYVQ 1378
 Db 1318 ATEVPVAVSQHPTKSRLOQSSLSLSARH-KAVEFPSSGAKSPSKSGAQTPKSPPEHYVQ 1377
 Qy 1379 ETLPMFSRCTSVSLDSFESRSTASSVQSEPCGMVSGIISPSDLPSDQGMPSRSKT 1438
 Db 1378 ETLPMFSRCTSVSLDSFESRSTASSVQSEPCGMVSGIISPSDLPSDQGMPSRSKT 1437
 Qy 1439 PPPPQTAQTKREVPKNKAPTAEKRESGPKQAANAQVQVLPDADTLHFATSTPD 1498
 Db 1438 PPPPQTAQTKREVPKNKAPTAEKRESGPKQAANAQVQVLPDADTLHFATSTPD 1497
 Qy 1499 GFSCSSLSLSLDEPFIQKVELRIMPVQENDNGNETESQPKSENQEKAEKID 1558
 Db 1498 GFSCSSLSLSLDEPFIQKVELRIMPVQENDNGNETESQPKSENQEKAEKID 1556
 Qy 1559 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKPAQATSKLPPPVARKPSQLPVYKLL 1618
 Db 1557 SEKDLDDSDDDIEILEECIIISAMPTKSRKAKPAQATSKLPPPVARKPSQLPVYKLL 1616
 Qy 1619 PSQNRLOPKHVSFTPGDDMPVYCVGPTPIPFSTATSLDSTIESPPNELAAGVGRRG 1678
 Db 1617 PAQNRLOPKHVSFTPGDDMPVYCVGPTPIPFSTATSLDSTIESPPNELAAGVGRRG 1676
 Qy 1679 AQSGEFKEKDTITEGRSTDEAGGTSVTPIELDDNKAEEGDILAEICINAMPKGSH 1738
 Db 1677 IQSGEFKEKDTITEGRSTDEAGGTSVTPIELDDNKAEEGDILAEICINAMPKGSH 1736
 Qy 1739 KPRVKKIMQVQOQASASAPNKNQDGGKKKPTSPVKPIFQNTYRTRVRKNADSKN 1798
 Db 1737 KPRVKKIMQVQOQASASAPNKNQDGGKKKPTSPVKPIFQNTYRTRVRKNADSKN 1796
 Qy 1799 LNAERFVSNKSKONLKNNSKDFNDKLPNNEDVRGSGFADSPHHYTPIEGTTCFSR 1858
 Db 1797 VNTTEFTSNKSKONLKNNSKDFNDKLPNNEDVRGSGFADSPHHYTPIEGTTCFSR 1856
 Qy 1859 NDSLSLDDDDDDVLSREKAEELRKAENKESAEKVTSHTLSNQSSQAQIAKQPI 1918
 Db 1857 NDSLSLDDDDDDVLSREKAEELRKAENKESAEKVTSHTLSNQSSQAQIAKQPI 1916
 Qy 1919 NRCQPKLQKSTFPQSSKQDIPDRGAATDEKLOFAIENTVPCPSHNSLSLSDIDQ 1978
 Db 1917 NRCQPKLQKSTFPQSSKQDIPDRGAATDEKLOFAIENTVPCPSHNSLSLSDIDQ 1976

Qy 1979 -NNKNEPIKETBPDSQGEPSKPQASGYAPKSFHVEDTPVCFPSRNSLSLSISDEDD 2037
 Db 1977 NNNKNEPIKETBPDSQGEPSKPQASGYAPKSFHVEDTPVCFPSRNSLSLSISDEDD 2036
 Qy 2038 LQBCISSAMPKKPKSRLKGDNEKHSRNNGGTIGEDLTLDLKDIOQPDSEHGLSPDSE 2097
 Db 2037 LQBCISSAMPKKPKSRLKGDNEKHSRNNGGTIGEDLTLDLKDIOQPDSEHGLSPDSE 2096
 Qy 2098 NFDWKAQEGANGSVSSLHQ-AAAAACLRSQAQSDSDSILSLKSGISLGSFPHLTPDQEE 2156
 Db 2097 NFDWKAQEGANGSVSSLHQ-AAAAACLRSQAQSDSDSILSLKSGISLGSFPHLTPDQEE 2156
 Qy 2157 KPFTSNKGPRLLKPEKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSSEISGQMKOP 2216
 Db 2157 KPFTSNKGPRLLKPEKSTLETKKIESEKGIKGGKVKYKSLITGKVRNSSEISGQMKOP 2216
 Qy 2217 LQANPISIRGRMTIHPGVNRSSSSTSPVSKGPPKLTTPASKSPSEGGTATTSPRGAKP 2276
 Db 2217 LQANPISIRGRMTIHPGVNRSSSSTSPVSKGPPKLTTPASKSPSEGGTATTSPRGAKP 2276
 Qy 2277 SVKSELSPVARQTSQIGSSKAPSRSGRSDSTSPRAQOPLSRPIQSPGRNSISPRNGI 2336
 Db 2277 SVKSELSPVARQTSQIGSSKAPSRSGRSDSTSPRAQOPLSRPIQSPGRNSISPRNGI 2336
 Qy 2337 SPPNKLQOLPRTSPSTASTKSGSGKGYVTSYSGRQMSQQNLTKQCLSKNASSIPRSES 2396
 Db 2337 SPPNKLQOLPRTSPSTASTKSGSGKGYVTSYSGRQMSQQNLTKQCLSKNASSIPRSES 2396
 Qy 2397 ASKGLNOMNGNANKVELSRMSSTKSGSESDRSEBPVLRQSTFIKEAPSPFLARKL 2456
 Db 2397 ASKGLNOMNGNANKVELSRMSSTKSGSESDRSEBPVLRQSTFIKEAPSPFLARKL 2456
 Qy 2457 BESAFSLSLSPSSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIE 2516
 Db 2457 BESAFSLSLSPSSRPASPTRSQATPVLSPSLPDMSLSTHSSVQAGGWRKLPNLSPTIE 2516
 Qy 2517 YNDGRPAKRDHIAKSHSESPSRPLPNNRSGTGWKREHSHKSSSLPRVSTWRTGSSSILSA 2576
 Db 2517 YNDGRPAKRDHIAKSHSESPSRPLPNNRSGTGWKREHSHKSSSLPRVSTWRTGSSSILSA 2576
 Qy 2577 SSESSEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNTNSQTVSSGATN 2636
 Db 2577 SSESSEKAKSEDEKHVNSISGTQSKENQVSAKGTWKIKENEFSPNTNSQTVSSGATN 2636
 Qy 2637 GAESKTLIYQMAFVAKXTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKSSSIND 2696
 Db 2637 GAESKTLIYQMAFVAKXTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKSSSIND 2696
 Qy 2697 SKDNQ---AKQNYGNGSVPMRTVGLNRLNSFTQVADPOKGTETKEGQNNPVPVSETNE 2753
 Db 2697 SKDNQ---AKQNYGNGSVPMRTVGLNRLNSFTQVADPOKGTETKEGQNNPVPVSETNE 2753
 Qy 2754 SSIIVERTPFSSSSSSKSHSPSGTVAARVTFNTPNPSRKSADSTARPSPQIPTPVNNNT 2813
 Db 2754 SSIIVERTPFSSSSSSKSHSPSGTVAARVTFNTPNPSRKSADSTARPSPQIPTPVNNNT 2813
 Qy 2814 KRRDSKTDSTESSGTQPKRHSGLYVTSV 2843
 Db 2814 KRRDSKTDSTESSGTQPKRHSGLYVTSV 2843
 Qy 2816 KRRDSKTDSTESSGTQPKRHSGLYVTSV 2845
 Db 2816 KRRDSKTDSTESSGTQPKRHSGLYVTSV 2845

RESULT 3

T30258

adenomatous polyposis coli protein 2 - mouse

N:Alternate names: APC2 protein

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000

C:Accession: T30258

R:van Es, J. H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Miles, A.; Kuipers, J.

Curr. Biol. 9: 105-108, 1999

A:Title: Identification of APC2, a homologue of the adenomatous polyposis coli tumour sui

A:Reference number: Z20796; MUID:99147086; PMID:10021369.

770 -----ENATKTSY----- 778
Db
QY 1110 SETNRVSGHGINQVSQLCOEDDYEDDKPTNYISRYSEBEEHBERPTNYSIKNEE 1169
Db
779 -----QETDL--DQPTDFSLYAE-----NQIESLDLISGPAQQ 811
QY 1170 KRHYDQPIDYSLKYATDIP--SSQKQSFSSKSSGQSKTEH-----MSSSENSTPSS 1223
Db
812 KSTITPP-----AETVPKSEGEILLILLDDSVKCYQIETDPYVLSNAASVTDLRVAA 864
QY 1224 NAKQONQLHPS--SAQSRGQPO---KAATKVSSINQETIQTYCVEDTPTICFSRCSLS 1278
Db
865 KADREAEVKPEVREVTSEKAPKLPKLSQCGSGSYTPKPIYCEBGTPTGYFSRYDLS 924
QY 1279 SLSSAEDEIG--CNQTOEADSANTLQAEIKETGTSARDPVSEFAVSOHPTKSSRL 1337
Db
925 SL-----DESKANQAIQVTD-----ADIKPLEKEBESQPAEQVLTKPPTQANSAL 973
QY 1338 QGSSLSSESARHKAVFSSGAKSPKSGAQTPKSPPEHYVOETPLMFSTRCTSVSSLDSE 1397
Db
974 -----ETPLMFSTRSSMDSLVHP 992
QY 1398 SRSTA-----SSVSEPCSGMVSGIISPSDLPDSGOTMPPSRSKTTPPPQACTKREV 1452
Db
993 DVDVANDCKSSVSD--FSRLASGVISPESEIPDSPTQSMPOS----- 1033
QY 1453 PKNKAPTAEKTESGPKOAAVNAQVQVLPDADTLHFATETSPDGFSSSSLSALSLD 1512
Db
1034 PRNSVAGSGQVNDSPVWIPASIQPLRSVFE--DDLSSFVETPTAPQFSTATSLSLSI- 1091
QY 1513 BPFTQKQVELRIMPVQENDNGNE-----TESHQKESNENQKE-AEKTID 1558
Db
1092 -----VDDEKAPAVMTEDNEDELLANCINMGQRKPTAVKSTVYNSEVDVAETIR 1144
QY 1559 S-----EKD-----LLDD--SDDDDI--- 1572
Db
1145 SYCTEDTPALLSKVPSNTLSVISMSTDPKDATAGQAQMAHQSLDDVSNASDCOGAS 1204
QY 1573 -EILEECIISAMPTKSRKAKKPAQTAASKLPPPVARK--PGOLPVYKLLPSONRLQPOKHV 1630
Db
1205 GHLLQOCTIRDM-----KKPLGEATSDPIAMLRGNGELPGY--LPS----- 1244
QY 1631 SFTPGDDMPRYCYVEGTPINSTATSLDITIES-----PPNELAAGVGRGQSGSEFEX 1686
Db
1245 -----ADENMK--FLVEDSPCNFVSGSLNLTGVGSLGPAVOLKETE--PSSADQNPENKR 1298
QY 1687 RDTIPTEGRSTDEAQGGKTSTVITPELDDNKABEGDILAEQ----- 1727
Db
1299 SLANRSKSRPPHQDDSLSLSDSEDDTNLLSQAIAAGCNRPKSNLGFSSNGKRSSSL 1358
QY 1728 -----INSAMPKGSHKPRVKINDQVOQASASSAPWQO-----LDG--KK 1769
Db
1359 SSSQPIAINAATSASSLSNMTVRKSOQESYSVSDSDSNQSKSLFELCILKGYKT 1418
QY 1770 KKP-----TSPVKPIPONTYEYR-----RVRKNADSKNLAER 1803
Db
1419 KEPGARAQMOEQPIVGSVSSVQSNPSLKQFDSLPVLQPSGQVQRQHHHHHHHRRER 1478
QY 1804 VFSNKKSK--KONLKN--SKDFNDKLPNNEDVRGSPAFSPHYTPIEGTPTCYFGRND 1860
Db
1479 ---ERKDEKLLQECINTGISKKIN-AVPKNV--LATSAALPECH--PMAAT-----T 1523
QY 1861 SLSSLDFFDDDDVLSREKAEARKAKENKESAKVTSHTELTSNQOSANKTOA-IAKOPI- 1918
Db
1524 SASALSTAAPDVE---QKAH---ATSNPQOQSSTHPSHILPNPIDAIVTDTARSPAA 1577
QY 1919 -NRGQPKPILOKQSTFPOSSKDIPDRGAATDEKLQNAFIEN-----TPVCF----- 1963
Db
1578 PNOGNGNASONGLET-ATGSKOLDSEDSDESQSFIMETWRLDGALETICISGASEK 1636
QY 1964 -----SHNSLSLSLSDIOENKNNKENEPKETEPPDS 1995
Db
1637 HKOPDLMLKSVRLTMEFVTSAEQLRSSSHHSS-----NSHKNNSSNNTWNTESTCPND 1691

RESULT 5

T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N/Alternate names: Hypothetical protein EG:49E4.1
C/Species: Drosophila melanogaster
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: T13564
R/Species: L.; Papagianakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A/Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A/Reference number: Z17689

A;Accession: T13564		Db	
A;Status: preliminary; translated from GB/EMBL/DDBU		1509	-----KSKEPSRRESIAESLKAESTKDEKSAAPPK-----EASRPGSVVSVESKDETEKS 1557
A;Molecule type: DNA		Qy	760 LEAELDAQHILSETFDNIDNLSPKASHRSKORHKOSLYGVDVFTNRRHDDNRSNFNTGNM 819
A;Residues: 1-5327 <SPA>		Db	1558 KEP-----SRRESIAESAKPIEFREVRSPSVI--DGIKDESAPKESRD-----1601
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1		Qy	820 TVLSFYLTNTVLPSSSSR--GSLDSSRSRSEKORSLELERIGLGNVHPA-----TENPGT 872
C;Genetics:		Db	1602 ---SP-----LASKEASRPESVLESVKDEBPXSTKESRRESVAESPKADSTKDEKSPLT 1652
A;Cross-references: FlyBase:FBgn0025392		Qy	873 SSXRGLOISITAAQIAKVMEVSAIHTSQEDRSGSGTTELHCVTDERNALRRSSAAHTHS 932
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1		Db	1653 SK-----DISPESAVENM-DAPFKETSRPESAVGS-----MKDSMKSEKPS 1694
A;Note: EG:4924.1		Qy	933 NTYNFTSESNRTCSMPYAKLEYKRSNDLSNVS---SSDGYGRGQMKPSIESYSED 989
C;Superfamily: Drosophila 576K microtubule-associated protein homolog		Db	1695 RRESVKDGAQSRSTSRPASVAESAKDGADLKLRSPESTTQSKAAGSIKDEKSPLEASE 1754
Query Match 4.5%; Score 649; DB 2; Length 5327;		Qy	990 DESKFCYGOYPADLAHAKIHSANMDD-----NOCGLDTPINYSLKYSDEQLNS-----1038
Best Local Similarity 18.2%; Pred. No. 3,1e-18;		Db	1755 EASRPASVAESVXDEAEKSEKSRRESVAESKPLPSKEASRPASVAESIKDEAEKSEES 1814
Matches 614; Conservative 521; Mismatches 1263; Indels 968; Gaps 129;		Qy	1039 ---GRQSPSQNERWARPKHI---TEDEIKQSEQ---RQSRNQSTTYP-----VYT 1079
15 ALKWMNSNLRQELE-DNSNHLTKLETEASNMKEVLKQLOGSIEDEAMASSGQIDLLERLK 73		Db	1815 RRESVAESKPLPSKEASRPASVAESIKDEAEKSEKSRRESVAESKPLPSKEASRPASVA 1874
819 AKKLQDLTASQELDAEKQRELDLKEE---QEVVREIEAVFSRDMKQEQQIKAEUR 874		Qy	1080 ESTDDKHLKTPHFGQECV-----SPYRSGANGSETNRVGSNHNQVNSQSLCQE 1132
74 EMLN-----DSSNPPGVK---LRKXMSLRSYSGREGSV-----SSRGECSPPVPMGSPFPR 121		Db	1875 ESIDEAEKSEK---SRRESVAESKPLPSKEASRPASVAESIKDEAEKSEKSRRESVAEK 1933
875 EMPAEGTGGENEPDEEERYLIIEKEVEQY--TEDSIVEQSSMTKE-----920		Qy	1133 D---DYEDDDTPNYSERVSEE-EQHEERERTNYSIKYNEEKHRVQDPIDYSLKYATDIP 1188
122 GFVNGSRESTGYLEELKERSLLADLDKEEKEKWYAAOLNLTAKRIDSLPLT-----175		Db	1934 SPLPSKEASRPASVAESIKDEAEKSEKSRRESVAESKPLPSKEASRP-----ASVA 1985
921 -----EEIQKHQR-----DSQSEKKEKSAEIEIAIAKVAEAEKARL 961		Qy	1189 SSQKQSFYFKSGSGQSKTEHMSSENSTSTPSSNA-----KRONLHPSSQAQRS 1240
176 ENFSLOTMTTRQLEYEARQIRVAMEEQIGTODMEK-RAQRRIAR-----220		Db	1986 ESIDEAEKSEKSRRESVAESKPLPSKEASRPASVAESIKDEAEKSEKSRRESVAES 2045
962 EGASARQDESELDEPEQSKIAEVQDIITATANDIAKSTRTEQLAKPAEELSSTPTEK 1021		Qy	1241 GQPKAATCKVS-----SINGETIOTYCEDTPICFSCSSLSLS--SAEDRIGC 1289
221 -----EIQLEK 226		Db	2046 PLPSKEASRPASVAESIKDEAEKSEKSRRESVAESKPLPSKEASRPASVAESVKDE--A 2103
1022 LSKKTSKDDQIGAPVDVLPVNLQESLPEEKSFATIESGATTAPTLPEDEIRIPDQIKE 1081		Qy	1290 NQTQEE-----ADSANTLQI---AEIKEKIGTRSAEDPVSEVPVAVSQHPRTKSRLOG 1339
227 DILRIQLQSQATEAB---RSSQKHQ-----TGSHDAERQNEQGVGEINM 271		Db	2104 DKSKESSRRESMAESGKAQSIKGDQPLKEVSRPESVAESVKDDPVKSEPSRRES--VA 2161
1082 DLVIEEKYVKEETKEAEIATVATQTLPEAAPLAIDTILASATKADKADANAALGE--L 1139		Qy	1340 SSLSESARHKAFFESGAKSPKS--CAQTPKPPPEHYQETPLMPSRCTSVSSLSOSFES 1398
272 ATSGNCGSTTRMDHETASVLSSSSTHAPRLTSHLGTKEVYVSL-----LSMLGTHDK 327		Db	2162 GSVTADSGARD-----QSPLESKASRPESVDSVDEAEKQSRRES-----2204
1140 PDSGE-----RVLPKMTFEAQNLRLRDVIKTPDEVADLPVHEEADLGLYEK 1186		Qy	1399 RSIASSVQSPFCGSMVSGIISPDLDP--SPGQWTPP--SRSKT-----PPPPQTAQTKR 1450
328 DMSRTILLAMSSQDSCISMROS-----GCLPLLIQLLHGNDKDS-----367		Db	2205 -----KTESVIPPKAKDDKSPKEVLQPVSMETIREDADQPMKPSQAESRR 2250
1187 DSODANAKSISHKEESAKEKETDDEKENKVGIEILGDEPNKVDISHVLLKESVQEAOK 1246		Qy	1451 E-----VPMK--KAPTAEKRESGPKQAA--VNAAVQVQVLPDADTLHLHATSTPD 1498
368 VLLGNRSGKEAPARASALHNIHQDPDKRGRREIRVLHLEQIRAYCETCWQEAH 427		Db	2251 ESTAESIKASPRDEKSPKASKEASRPESVAESIKYDLDKPKIHKDDKSTEHRSRESLED 2310
1247 VVIETTVKQBEIATEATVITQENQED-----LMEQVKD-----KEEH 1286		Qy	1499 GFSCSSLS-----1507
428 EPQMD-----ODKPMFPAPVEH-----QICPAVCLMKLSFDEEHRHAMN 467		Db	2311 KSAVTSEKSVSRPLSVASDHEAAVAIEDDDAKSSISPKDKSRPGFVAETVSSPIEETMEF 2370
1287 EQIESGIITEKEAKSASTPEKETSDITSDELPQAQADPTTVPKSAKDR-----1340		Qy	1508 -----ALSLD-----EPFIQKDELIRIM 1525
468 ELGGLOA-----IALLQVDCMYGLTNDHYSITLRRYAGMALTNLTFGDVANKATICSMMKG 524		Db	2371 SKIEVVEKSLALSLOGSGGKLQTDSPVDVAEGDFSHAVASVSTVTTLTTPKPAELAQI 2430
1341 DTGSIESPTIEAIEVE-----1358		Qy	1526 -----PPVOENDNGNETESEQPKSENQOEKAEKTIIDSEKDLDDSD 1569
525 CMRALVAQLKSESED:QQVIASVLRLNLSMRADVNSKTLREVGSKALMECALEVK-KES 583		Db	2431 GAAKTVSSPLDEALRTPSAPEHISRADSPACASEIASQDKSPQVLKSSSRPAWVAESK 2490
1359 -----VQAKQEAQ--KPVAPAEALIKTEKSPKASKEATSRPESATGSKVEDTEQTKSKKS 1410		Qy	1570 DDTEILEECIISAMPTKSSSKAKKPAQATSKLPPPVARKPSQPLVYKL-----1617
584 TLKSVLSALNLSAHTENKADICAVDGAFLVGLTYRSQTNTLAIIESGGGILRVNS 643		Db	2491 DDAQLKSSVEDLRSPVASTEIRSPASAGETASSPIEAPKDFABEQAEKAVLPTLIEL 2550
1411 PVPS-----RPEAKDKKSPASGEAS--RPESVAESVKDEAKAESRRE 1454		Qy	
644 SLIATNEDHRQI---LRENNCOTLLOHLKSLTIVSNACGLTNLNLNLSARNPKDQEAALWD 700		Db	
1455 STAKTHKSSSLDKAKEQSRRESLAESIKPSGGIDSKSA-----LASKAESRPESVTD 1508		Qy	
701 MGAVMSLKNLHLSKHKRVIMAG--SAAALRNLMANRRPAKYKDNANINGPSSLSPLSHYRKOKA 759		Db	

1668	QY	ELAAEGVYRGGAQSGEPEKRDIP--TEGSTDEAOGGKTSSVTIPE-----LDDNKAE	1720	QY	-----LDDNKAE
1081	DB	--TEASSTGGSTSPNPGQSTSPSTGATSPGSGSGTTLTSLSPSPQSSITGSSQGST	1138	DB	
1721	QY	GDILAEINAMPKPKSHKPRFKIMQDVQOQASASSAPNKNQLDKKKKTPSPVKPIP	1780	QY	
1139	DB	SPVSTSGDMTSQGSTQIPGSTGTVTQPGSTGSGSTSGEITSGQSTQTPRSSJLSTSP	1198	DB	
1781	QY	QNTETRTVRKNADSKNWLNAEVPNDKNKSKONLKNNSKDFNDKLNNEDRVRGSFAF	1840	QY	
1199	DB	AIS-----TSTQSVSTNSPGSTVTQPS-----VRGSTSS	1229	DB	
1841	QY	DSPHPTIEGTPYCFSRNDSLSSLDFFDDDDVLSREKAEALRAKENKESAKVTSHTL	1900	QY	
1230	DB	GSTVTGTSGE-----STSGSSSATSLSSSPV	1258	DB	
1901	QY	TNQOSANKTQIAKQPINRGOPKPILOKQSTFPQSSKDIIPRGAATDEKLQNFALENTP	1960	QY	
1259	DB	PSTOSPNPSTSGSSTP-----TFNP-----SQSTSPVVS-----TTTGBMTSHGSTQTP	1303	DB	
1961	QY	VCFSHNSLSLSDIDQENNNKENEPKITEPPDSQCE-PSKPOAGYAPKPSHVEDTPV	2019	QY	
1304	DB	STI-----GSTVTQSTVSGSNSGSGVTIGSEASTSGSSPKTSPSSISPVPTSPPISTTF	1361	DB	
2020	QY	CFSRN--SSLSSLSIDSDDLQECISSAMPKKKXKPKRLKGDNEKHSFPRNMGWILGEDJTL	2078	QY	
1362	DB	ASSTSGSTISDVSSVSTTSLAP--JSSSLPSTV-----	1392	DB	
2079	QY	DLKDIDRPPSEHGLSPDSENFWDKATQEGANSIVSSLHQAAAACLSRQASSDSDSILSL	2138	QY	
1393	DB	-----PSTOSFSST-----EGSSKASSPVPSQSTSTPNPTGSTTESSTL-L	1435	DB	
2139	QY	KGSI LSGSPFHILTPDQ-----EBKPFSTNKGPRILKPKCKSTLTETKIESEKGIKGGKV	2194	QY	
1436	DB	SSTIS--GSTCHITMSKASSGSTSPSNTSGSTVITWGSST-----SGV-----	1478	DB	
2195	QY	YKSLITGKVRNSIEIGQMKOPIQANMPSISRGRTWMIHIPGVNNSSTSPVSKKGPPLK	2254	QY	
1479	DB	-----STSGASTQPMQSTSQSSAGSTV-----ASTAGPAA-----	1511	DB	
2255	QY	TPASKSPSEQATTSPRGAKPVKSELSPVARTSQIGSSKAPSGSRGRDSTPSRPAQ	2314	QY	
1512	DB	--SSTAPSSTGTMSTSGTSGVTISESSTTASASQSTGVTWGSSTSGVSTSSAST	1569	DB	
2315	QY	QPLSRPIQPGRNISIPGRNGISPPNKLQOLPTTSSPSTA-----STKSGSGKMKVTS	2369	QY	
1570	DB	QP-----QWSTSQG-----SSAGSTVASSTAGLVSTSTVPSTGTMTGSTSS	1610	DB	
2370	QY	GRQMSQNLTKOTGLSKNASSIPRSESASGLNMNNGCANKVELSRMSTKSSCGSES	2429	QY	
1611	DB	-----TWGSTISESSTTASASQSG-STVTWGSSTSGVSTSSASTQPMQSTS	1659	DB	
2430	QY	DRSERPVLQRST-----FIKEAPSPILRRKLKESAFESLSPSRPASPTRSQATPVLS	2485	QY	
1660	DB	QGSAGSTVASTTGLVSTSTVPSTGTMTGSTSGTSGVTISESSTTASASSQSTGVTM	1719	DB	
2486	QY	PSLPDMSLSTHSSVQAGWRKLPNNLSPTTEYNDGRPAKHDDIARSHSESPRLPINRSG	2545	QY	
1720	DB	GSSSTSGVST--SSAGSG-----QPMQSTSQGSAGSTVTVSSTASPAASTAPSTGT	1769	DB	
2546	QY	TWREH-----SKHSSSLPRVSTWRRTGTSSSSTLSASSESSEKAKSDEKHNVISGTHQ	2600	QY	
1770	DB	TMSSTSGTGVSTGSSQSSSTAASSTHSGSTVITLGSSTSSNQWSTSGSSVGSVASTA	1829	DB	
2601	QY	---SKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTLIYQWAPAVSKTEDV	2657	QY	
1830	DB	GLVSTSTVPSTGTMTGSTSGTGVSTISESSTTASASSQSTGVTM-----	1875	DB	
2658	QY	WVRIEDCPINNPRSGSPGTCNTPPVDSVEKAPNPKDKSKDNCAKQNVGNSVPMRTVG	2717	QY	
1876	DB	-----GSSSTSG-----VSTSSASTQPMQSTQGSAGSTVASS-----TAG	1913	DB	
2718	QY	LENRLANSFIQVDAPDQKGTGIKPCQNNPVPVSET-NESSIVERTPFSSSSSKHSSPSGT	2776	QY	

Db 1914 LVSTST-----VPSSTGT---NGSTSGTVGTSISSSTAAGTSSQTGTVTIGTSGT 1964
 QY 2777 VAA---RVTPNNYNSPKRSASDTSARPQISQITPTVNNNTKKDSKTDST-----2823
 Db 1965 NPSPRLSLQITITPSSQSTESTQTSLSPPSSPSTHSVSSSGTMSGATTSGDKMS 2024
 QY 2824 --ESSGTQSPKRHSGLVTS 2842
 Db 2025 FLSGTGTVSFSSRGSLATT 2045

RESULT 8

T23327
 adenomatous polyposis coli protein 1 - Caenorhabditis elegans

N;Alternate names: apr-1 protein
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C;Accession: T23327; T03822

R;Gardner, A.

submitted to the EMBL Data Library, July 1996

A;Reference number: Z19727

A;Accession: T23327

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1186 <WIL>

A;Cross-references: EMBL:Z19727; PIDN: CAB00045.1; GSPDB: GN00019; CBSP: K04G2.8a

A;Experimental source: clone K04G2

R;Rochelleau, C.E.; Downs, W.D.; Lin, R.; Wittmann, C.; Bei, Y.; Cha, Y.H.; Ali, M.; Prie

Cell 90, 707-716, 1997

A;Title: Wnt signaling and an APC related gene specify endoderm in early C. elegans embr

A;Reference number: Z15051; MUID: 97433081; PMID: 9288750

A;Accession: T03822

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1186 <RQC>

A;Cross-references: EMBL:AF013950; NID: g2338717; PIDN: AAC47747.1; PID: g2338718

A;Experimental source: strain Bristol

C;Genetics:

A;Gene: apr-1; CBSP: K04G2.8a

A;Map position: 1

A;Introns: 61/3; 113/2; 355/3; 551/3; 1067/3; 1161/3

A;Note: apr-1

Query Match 3.5%; Score 511.5; DB 2; Length 1186;
 Best Local Similarity 22.0%; Pred. No. 1.8e-13;
 Matches 286; Conservative 217; Mismatches 580; Indels 219; Gaps 50;

QY 271 MATSGGQGTTRMDHETASVLSSTHSAAPRLTSHLGTQVEMVYSLLSMLGTHDKDM 330

Db 1 MSSSSDENETT--IHRGNTGGGIYQPRAGSGKRTSNVRHDVS-----DVDDE 50

QY 331 SRILLAMSSQDSCINRQSGCLPLLIQLLHGNDKDSVLLGNRSGSKEARASAAALHN- 389

Db 51 EEHYARF--REDTAIEV--DDAITVLSLHFHKKRDIPTDEDDNK-----LRELHEK 100

QY 390 ---IHSQPDKRGREIRVHLLEQIRAYCETCEWQAHEPGMDQKNPAPVEHQI 446

Db 101 IFALITSESVDNRKRLKALPASNCVR-----EQVYV--LRKKESTPPASYHEL 149

QY 447 CPACVLMKLSFBEHRHAMVELGGLOAIAELLQVDCMYGL--TNDHYSITLRRYAGMA 504

Db 150 NAALHTIVKESFGEEYRKVATVGLVLEALAEVLEVHTFGINETPGSHRNIRKLIANA 209

QY 505 LTNLTGFDVANKATLCMKGCMKALVAQLKSEBEDIQQVIASVLRNLNRADVNSKTLR 564

Db 210 LTNLTGQIHSKRLCSYDGFIRCVV--RVIESPNITQVAGLRNLNSWADSGSEALQ 268

QY 565 EVGSGVAMCEALEVKES--TLKSVLSALWNLSAHCTENKADICAVDGAFLVGTLYR 623

Db 269 P--TVHALSIAAVHAHTHFRDVTATLSALWNLGASHVENKRTICDTPNCKLVLASLLSPD 326

QY 624 SQNTLAIIESGGILRNVSLLIATNEDHQILRENNCLQTLHLKSHSLIIVSNACGT 683

Db 327 ARFTSL--VDSATGILKVVSOVLANTSTHLEL--RSLILTRMLLLKSASFTCVNTLGA 382
 QY 684 LWNLSARNPKDOAL--WDMGAVSMLKXLIHSHKXKVIAMGSAALRNLMANRPAYKANI 742
 Db 383 TANLTVKDPHQWIRQDMAAVQQLNLVLRNRRDITAVKSVLNTL--NQPCSHRYGDM 440
 QY 743 M--SPGSSPLSLHVRKQKALEAEALDAQLSETFDNMLSPKASHRSQRKHQSLYGVYVF 801
 Db 441 SHSVGGGATGMQLSEPOLQMQ--TSHHAYHGTASPRLLSLRATRASPGKIQAQOQLI 499
 QY 802 DTNRHDDNRSD---NFNTGNMTVLSPYINTTVLPSSSSSSRGSLDSSREK--DRSLEREG 857
 Db 500 QTPQVDQRSSSLPRHF-----AVQRNGFVMAQSYNQ--MDHQOQQQMIYLOQOQQ 549
 QY 858 IGLGNYHPATENPGTSSKRGQLQISTTAAQIAKVMEEVSAIHTSOEDRSSGTTTELHCV-- 915
 Db 550 I-----MTEDQAMEHH-----QQIMYLOQOQQQFHOIQOQQOMKQAEADPVPP 594
 QY 916 TERNALRRSSAHTHTNTYNTKSENRTCSMPYAKLEYKSSNDSLNSVSSSDGYGK 975
 Db 595 TDDLDIPTSTVMGTRNSRSLGSMNPGSVMTWNSSLDTAANSSRALSPVSYND--IPA 653
 QY 976 RGQMKPSIESYSEDESKFCSYGOVPADLAHKIH---SANHMDNDGELDTPIN----- 1026
 Db 654 SPTCAQVFNLPKTESE---HHOLTQQQNTTHYSSGSANTMTSDGATVPMDNITP 710
 QY 1027 -YSLK---YSDEQLNSGRQSPQSNRWARPKHIEDEIKQSEQ-----RQSRNQ 1072
 Db 711 TVAILNPLVHEQTPNGTPVPRKTSSELDPSDDVLPGPSLEEEDGYAIIIGAAKTDDEL 770
 QY 1073 TTYPVYVE--STDDKHLKQPHFG-----QCECVSPYRSGANGSETNRV--GSNHI 1121
 Db 771 LTRSIQSEMTSSTPKMKVSPRLNGFFPTQKTTSPAWSH-----PDTSPKSSSHRT 826
 QY 1122 NQVYSQSLQCEDDY-----EDDKPTNYSERYSEREEQHEEERPTNYIKYNEEKHVD 1174
 Db 827 QPNRRQDASDADRLMESIMSEMPKRIISPRLAGTQOYLEPE--PERRSHSKNEEADRD 885
 QY 1175 -----QPIDYS--LKYATDIPSSQKQSFPSKSSGSSGSKTEHSSSENTSTPSSNAKR 1227
 Db 886 AFTASHEPDSHNGIDVARGSDWSPQQLHRMESLESQASSEDSFGLTAEPNSSTGSA-- 943
 QY 1228 NQLHPSSAQSRSQPKAATCKVSSINQETIQYCEVEDTICFSRCSLSLSLSAEDAI 1287
 Db 944 -----AANTMRDDEIDASLPMDCVDDDDYD---TYDHPEDYEDEE 982
 QY 1288 GCNTOQASANTLOIAIEKEKIGTSAEDPVEPAVQHPRTKSSRLQGSLSSES 1347
 Db 983 DPDAT--QFDGVDQAQLTIDCSMISSGSSGSRNETTTTTRDSKALATSTPKGSASSLP 1040
 QY 1348 RHKAVERSSGAKS---PSKSG-----AQTEKSP-----EHYVQETPLMF 1384
 Db 1041 VRQATRVSTNGKSLPVKTINGSLVDKNPKPIIASRPRPLPPKPTLLKXHYPEE----- 1095
 QY 1385 SRCTSVSLDSFESRS-----IASSVQSEPCSGMVGSIISPSDLPDP--QGTWPPSR 1435
 Db 1096 -----DSIENQTRDDTIYVNAVPVVEAEQERIVYNNALKQKNIQBSPSIGNSPIAK 1146
 QY 1436 SKTPPP-----PPQATQTKREVKNKAPTAEKESGPKQAAV 1472
 Db 1147 SAIVTPYNYQKPPPTGRNNGEMSEKSVT-----PNPKQMLV 1183

RESULT 9

T23330

hypothetical protein K04G2.8b - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T23330

R;Gardner, A.

submitted to the EMBL Data Library, July 1996

A;Reference number: Z19727

```

773 LRFISIQSENPSSSTPKMKVSPRLNGFFPTQKTTSPPAWSH-----PDTSPIPKSSSHT 828
1122 NQNVSQLCQEDDY-----EDDKPTNYSERYSEEEQHEEBERTNYSIKYNEEKHYVD 1174
829 QPNRRQASDADRLMESINSEMPKRIISPLAGTQQYLEPE-PERRSHKNEEAADRD 887
1175 -----OPTDYS-LKVATDIPSGOKSFSGKSSSQSKTEHMSSESSTSTPSSNAKR 1227
888 AFTASHEPDHNGIDVARGSDNSPQOLHMESELESQASSEDSFGLTAEFNSSTSGA-- 945
1228 QNQLHPSSAQSRGGQPKAATCKVSSINQETQTYCVEDTPTCFRCSSLSLSSAEDEI 1287
946 -----AANTWFDDEIDASLPMDCVDDDDVDY----TVDFHPEYDEE 984
1288 GCNQTTQEADSANTLOIAEIKKIGTRSAEDPVSEVPVAVSQHPRTKSSRLQGSLSSESA 1347
985 DDPAT--QPDGVDQAQTIDCSMISSGSGSQORNETTTSRDSKALATSTPKSASSLPG 1042
1348 RHKAVEFFSGAKS-----PKSG-----AQTPKSP-----EHYVOETPLMF 1384
1043 VRQATRVSTNGKSRLPVPKTNGSLVDKNPKPIIASRRPLRPKPKPTLLKDKHYPEE---- 1097
1385 SRTCSVSSLDSPERS-----IASSVQSEPCSGMVSGIISPSDLPDSP--GQTWPPSR 1435
1098 -----DSINQTRDDPIYNAPVVEAQERIYNALKQOKNIEQSPSINGSGPIAK 1148
1436 SKTPPP-----PQTAQTKEVPKNAKPTAKRESGPKQAAV 1472
1149 SAIVTPYNYQKPPFTGRNNGEMSEKSVT-----PNPKQMLV 1185

RESULT 10
A43359
microtubule-associated protein MAP1A - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Aug-1999
C:Accession: A43359; S22108
R:Langkopf, A.; Hammarback, J.A.; Muller, R.; Vallee, R.B.; Garner, C.C.
J. Biol. Chem. 267, 16561-16566, 1992
A:Title: Microtubule-associated proteins 1A and LC2. Two proteins encoded in one
A:Reference number: A43359; MUID:92355629; PMID:1379599
A:Accession: A43359
A:Molecule type: mRNA
A:Residues: 1-2774 <BLN>
A:Cross-references: GB:M83196; NID:G205937; PIDN:AAB48069.1; PID:G205538
A:Note: Sequence extracted from NCBI backbone (NCBIN:111039, NCBI:P:111040)
R:Cravchik, A.
submitted to the EMBL Data Library, June 1992
A:Reference number: S22108
A:Accession: S22108
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 73-364,'NRLRS',370,'QKN',374,'PSPKGL',381-751,'RSMGQNNLQRR',764,'D',
'WLKENMCPQRPSP',851,'V',853,'NSL',855,'LPHRWLRN',865,'W',867,'HSQLPDGDD',877,'
A:Cross-references: EMBL:X66840
A:Experimental source: strain Sprague Dawley
C:Superfamily: microtubule-associated protein MAP1B
C:Keywords: microtubule binding; phosphoprotein

Query Match 3.3%; Score 476; DB 2; Length 2774;
Best Local Similarity 18.6%; Pred. No.1.6e-11;
Matches 540; Conservative 364; Mismatches 1038; Indels 954; Gaps 130;

QY 609 VDGALAFIVGLTLYRSQT-----NTLAIE--SGGGIL-----RNVSLLIATN 649
DB 1 MDGVAEP-----SEVSVETDVPSPFDLLEPTSGFLKLPCCYIPFGRGDSALFAVN 56
QY 650 -----EDHRQILRENNCLQTLOHLKSHSLTIYSNACGTLWNLSARNPKDOEALWD 700
DB 57 GFNIIVDGGSDRK-----SCFWKLVRHL-----DRIDSVLLTH 89
QY 701 MGAVSM--LKNLIHSKKHVIANGSAALRNLMANRPAYKD--ANIKSPGSSLSLHV-- 754

```

Db 90 IGADNLPINGLLQK-----VALEBEQSGSSYSYDWMVKNLISPGLVVFVFNVPD 141
 Qy 755 -----RKOKALEAEADAOHLSETFDNINLSPKASHRSQR-----HKQSLYGDYVF 801
 Db 142 KLRLPDASRAKASIEACLTQHILNRLGLOAEPLFRVSVNTIEPLTLPHKMGVGRLDWY 201
 Qy 802 DTRHDDNRSDF-----NGNMVTLSPYLNTTV-----LPSSSSSR 838
 Db 202 VLPNVKDSKEMQFLMQKWNAGSKAKTIGVLNKGAEISVPYLSITLWVLPANPTEK 261
 Qy 839 -----GSLDSSR-----SEKDSLE-----REGIGLG-----NYHATENPFTSSKR 876
 Db 262 IVRVLPFNAPQNKILBGLKRLHLOFLRYPVATQKDLAAGAVPANKFSKIKHRADSKE 321
 Qy 877 GLOIS-----TTAAQIAK-----VMEEVSIAHTSOEDRSSGTTTELHCVTDERNALRRSSAAHTH 931
 Db 322 SLKAAAPTAVSKLAKREEVLEEGAKARSSELAKEKAKPEKPEKPEKSERVR 381
 Qy 932 SNTYNTKSEN-----SNRTCSMPYAKLEYKRSS-----NDSLSNVSSSDGY-----973
 Db 382 GESSEALAKERLLIKOKAGKHLKEKISKLEEKOKKEKEIKERKELKTEEGRKEEK 441
 Qy 974 -----GKRGOMKPSIBSDESKFCVQYQPADLAHKIHSANHMDNDGELDTPINYS 1028
 Db 442 DAKDEKGRKDTKEVKKLSKEDLKPF-----TPVRKTLYKAKAPGRVYDKGPA 491
 Qy 1029 LKYSDEQLNGRQSPQNE-----RWAPKHIIEB-----BIKOSORQRNOST 1073
 Db 492 ARGEKELSEPRTPPAQKGAAPPAAGHRELALSSPEDLTQDFELKREERGLLAEQD 551
 Qy 1074 T-----XPVYTESTDDHLK-----FOP-----HFGQECVSPVRSGANGSETNR-- 1114
 Db 552 TGLGEPLPADATEQHPNSALIVTQSPGVLEGEHVEREKEVVP--DSPEDKGS--TNRGP 609
 Qy 1115 -----VGSNHGINQNVQSLOQEDDYEDDKPTNYGERSSEEFQ 1152
 Db 610 DSGAEVEKEKETWEERKQREAEELGPENTAAREBEAEVKEVIEKAELEMEETHPSDEE 669
 Qy 1153 HEEERPTNYSIKYNEEKHVQPIDYSLKYATDI-----PSSQKQSFSSKSSG 1203
 Db 670 GETKAESEFY-----QXHTQALKASPKSRAUGRDLGFGKAKPEKETASFUSSLAT 722
 Qy 1204 QSSKTEHMSSESSSENTSTPSSNAKR-----ONQLH-----PSSAQSRSGQ-----PQ 1244
 Db 723 PACATEHVSVIQDETIPYSETEQITISDEEIHDEPDERPAPPRFTSTYDLSGSPGPP 782
 Qy 1245 KATCKVSSINQETIQTYCEDTIPCFRCSLSSLSAEDIEIGNQTTQOADSANTLOI 1304
 Db 783 EASQAADSAPVASSSKTYGAPETELTY--PNMVAAPLAEER-----HVSSATSITE 832
 Qy 1305 AEIKKEIGTRASBD--PVSEVPVAVSQHPRTKSRLOGLSSLSARSARKAVEF-----SSGA 1358
 Db 833 CDKLSLSFATSVADQSVASLTAPQTEETGKSLLLDVTISIPSSRTEATQGLDYVPSAGT 892
 Qy 1359 KPSKSGAQT--KSPPEHYVQETPLMFRCSYVSLSDFS-----SRSIASS-----1404
 Db 893 ISPTSSLEEDKGFKSP-----CEDFSVTGESEKKGTGVRGLSGERKAVGKE 939
 Qy 1405 -----VOSEPCSMVSGIISPSDLPSPGOTMPPS-----RSKTPPP 1441
 Db 940 EKVVVTISEKLSGOYAVFG-----APGHTLPPGPALGEVEERCLSPDDSTVQKASPPP 993
 Qy 1442 --PQTAQTKREVPKNAFTABKE-----SGPKQA 1470
 Db 994 SGPPPSAAHT-----PFHQSPVEDKSEPRDFQEDSWGETKHS PGVSKEDGEEQTVKPGPEG 1049
 Qy 1471 -----AVNAVQ-----RVQVLPDAATLLHFATFESTDPFGSCSSLSA 1508
 Db 1050 TSEEGKGPTRPQADQMPVSIAGGTGCTIQLLPEQDAIVFTG-----EAGNLGA 1103
 Qy 1509 LSL-----DEPFTQKQVLR-----IMPPVQENDNGNETESEQP 1542
 Db 1104 GTLPGEVRTSTEEATEP--QKDEVLRFTDQSLSPEDAELSLSVSVSPDTTKQATPRSP 1161

Qy 1543 KESNENOEKAEKTIIDSEKDL---LDDSDDDDIIEILECIISAMPTK-----SSRKAKK 1593
 Db 1162 CSLKEQO-----PHKDLWPMVSPEDTQSLSFSE-----SPSKETSLDISSKQLSP 1207
 Qy 1594 PAQTASKLPPPVARKPQLPVYKL-----LPSQNRLOPQKHVSFTPGDMP--- 1639
 Db 1208 ESLGTLOFGLNLGKEERGPWMAEADDSDCHLAPVSIPEPHRATVSPSTDETACTPAGGS 1367
 Qy 1640 -----RVYC-----VEGTPINPSTATSLDLTIESPPNELAA-----GEGVRGG 1678
 Db 1268 FSHSALSVDKXHSGEITPGGHEMT-----SDSLTKSPESLSPPAMEDLAVWEGKAPG 1323
 Qy 1679 AQSGEFKRDY-----IP-----TEGRSTDEAGGKTSTVTIPELDDNK 1717
 Db 1324 KEKEPELKSETROQKQILPEKVAVVEQDLI IHQDQALDEENKPGRQDKT--PEQKGRD 1382
 Qy 1718 ABECDILAECINSAMPKSHKPRVKKINDVQOASASSAPNKNQLDGKKKPTSPVK 1777
 Db 1383 LDEKDTAAELDKGPEPK-----EKDLREDQGORAGPPAKKAKASEQDRTLQOT- 1432
 Qy 1778 PIPONTEYRVRKRVNADSKNNLNAERVFSNKKQKQNLKXNKDFNDKLPNNEDRVRS 1837
 Db 1433 ---QATEPRDAQERRD-----SEKD--KSLELADRTPEEKDRILVQEDR--- 1473
 Qy 1838 FAFSPHHYPIEGTPYCFGRNDSLSLDFDDDDVDLSREKAEKLRKAE---NKESEAKV 1894
 Db 1474 ---APEHSIP---EPTQTDRAFDKGTDDKEQKEEASEEKEQVLEQKQWALKEGET-- 1524
 Qy 1895 TSHTELSNQOSANKTQAIKOPINRGPKPIQKOSTFPQSSKDIPDRGAA---TDEK 1950
 Db 1525 -----LQOEARTAEQDETILKEKTKQCKSSFVEDKTT--TSKETVLDDQSAKADSVQ 1577
 Qy 1951 LQNAFIENTPVCFSHNSLSLSLSDIOENNN--KENBPI---KETEPPDSQGP------S 2000
 Db 1578 QDGAALKETKALGLEESPAGSKAREQEKYKWEQDVVQGWRETSP--TRGEPVGQKEP 1635
 Qy 2001 KPQASGVAPKS--PHVEDTVPVCFSRNSSLSSLSISDSEDDLQECI--SSAMPK--KKKPSR 2055
 Db 1636 VPWEGKSPQEVYWRDRDITLQDAYWRELCDKRVWPPHELDGQGARPRYCEEREST 1695
 Qy 2056 L--KGDNE-----KXSPRNMGILGEDLTLDLKDQRPDSEHGLSPDSENFQWKAIEQ 2107
 Db 1696 FLDEGPEDEITPLQHTFRS-----PWTSDFKDFQEPQLPKGLEVER---WLA--- 1740
 Qy 2108 ANSIVSSLHQAANAACLRSQASDSDSILSKSGISLGSFPHLTPQEEK---PF----- 2159
 Db 1741 -----ESPVGLPPEEDKLTSPFEIIS 1763
 Qy 2160 -----TSNKGPRILKPOEKS---TLETKKIESK----- 2186
 Db 1764 PPASPPMTQVRPS--APQESPVDPDTSTAPRNEPTTPSWLAIEIPWVPKDRPLPPA 1821
 Qy 2187 -----GKGGKVVYKSLITKGVRSNSEISGOMQKPLQANMPSISRG 2227
 Db 1822 PLSPAPAPPTPAPPEHTPVFPFWSGLAEDSVVAVQEAELGEGYPSPGLKDYKRAEGE 1881
 Qy 2228 RTMHIHPGVNRNSSSTSPVSKGPPLKT-----PASKSPSE----- 2263
 Db 1882 REBEGGAGADSSSFPKVPFEGESLATRDTEQTEPQREPTPYPDERSFQVADIEQOM 1941
 Qy 2264 ---GQATTT--SPRGAK-----PSVKSELSFVARQTSQIGGSKAPSRSGSRDSTP--- 2309
 Db 1942 LTGLGPACPTREPLPGASGDWPPHLSKTEBAAGCNTSAEKETSSPASPNLQSDTAPSV 2001
 Qy 2310 ---SRPAQQLSPRIOSGRN---SLSPG-----RNGISPPNKLSQLPRTSSPT 2353
 Db 2002 ASLAGAVPPRQBP--DPGNVFPSTPFAVPPAPFISLKSOLSPPLNGSTVSCSPDRIT 2059
 Qy 2354 ASTKSGSGRMVSTSPGRQMSQQNLTKTGLSKNASSIPRSESASKLQ---MNGNGA 2410
 Db 2060 PSPKETGRGHW-----DDGTNDSLEKAREQPEKETRSPSPHPMPMGHSLW 2108

Qy	2411	NKVELSRMSTKSSGS-----	ESP	2430
Db	2109	PETAYSLSDSHLGSVRSLDPPAGAFGSSLOPAPPAPRSPAPCGSLAFSGD		2168
Qy	2431	RS-----ERPVLVRQSTFIKEAPSTLRRKLEESASFES-----LSPSSRPASP-----T		2475
Db	2169	RALALVPGTPTRTDHYLVETKAPSLDSSLPOLPSPSSPGGLNLNLPAPSPALSEGS		2228
Qy	2476	RSQQTPLVPSL-----PDMSLTHSSVQAG-----WRKLPNLSPTIYNDGR		2522
Db	2229	SSEATTIVISVAERFPFGLAEAFQSAEGLGSGKESAAHSLDLTP--LSPA-----P		2279
Qy	2523	AKRDIARSHSESPRLPINRSGTWKREHKGHSSLP-----RVSTWRRTGSSS		2571
Db	2280	SASLDLAPAPAPAPAP-----GLPGLDGDGTLPCRPECTGELTKKPS		2323
Qy	2572	SILSASSESEKAKSEDEKHVNS---LSGTQSKENQVNS---AXGTWRKIKENEFPTN-		2624
Db	2324	PFLSPSGD--HEANGPGETSLNPPGFVTATAKEEAEPHAWERGSWPEGAERSRPDTL		2381
Qy	2625	-STSGTVSSGATNCAESKTLIYQ-----APAVSKTEDVWV		2659
Db	2382	LSSQPIAPGKSSGPPCSSLSEVEAGPGCATDPRPHCGELSPFLNPLPSTDDSDL		2441
Qy	2660	RIEDCPI-----NNPRSGR-SPTGNTPPVIDSVEKAMPNIKDKONAKQNVNGSVPMR		2714
Db	2442	STEARLAGKGRRRVGRPGATGCPMAD---ETPPTASDSSGSSQSDSDVPPEBEC		2498
Qy	2715	TVGLENLNS-----FIQVD-APQKGT-ETKPGONNV-----PV		2748
Db	2499	SITAEALDSDGDFLVDKAGVSGTHHPRPHDPPPTPLDPRSPRPDPVCMADPE		2558
Qy	2749	SETNESSIVERTPFS-----SSSSSHSSPSGTVAAARVTPFNPNPSRKSADSTS---		2799
Db	2559	GLSSGSRVLEKRGPRRAPGRAPKAPASPARLDIR---GKRSPTGKGPVDRTSRTV		2615
Qy	2800	ARPSQIPTPVNNYTK		2815
Db	2616	PRPRSTPSQVTSABEK		2631

RESULT 11

hypotheical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favella, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 3.1%; Score 454; DB 2; Length 3507;
Best Local Similarity 19.9%; Pred. No. 1.7e-10;
Matches 465; Conservative 277; Mismatches 849; Indels 740; Gaps 100;

Qy 1051 RPKHIIEIRIKSEORSRNQST-----TYPVYTESTDDKHLKFOHFQGCQECV 1099
Db 629 RSNHSTSDLIQTQVQQRNFSTGQILLTRGVSSGEAVTQTDADEF----- 676
Qy 1100 SPYRSRANGSETNRVGSNHGNQNVQSLOEDDYEDDKPTNYSERVSEEBEHEE----- 1155

Db	677	-----GLEISAADLAGSGSGITLPTT-----LEPKIEGSKKASGVWTEDEGEDEDLM		726
Qy	1156	BERPTNYSIKYN-----EKRHVQDP-----		1176
Db	727	EEGSGSWSTTINGTGITGSPSEGTIRVIRITLGEDGEPETATKPGISAPDKTGEKSXTE		786
Qy	1177	IDYSLKYATDIPSSQKQSFPSKSSGQSSKTEHMSSESSENTTSPSSNAKRONQLHPSSA		1236
Db	787	SDGEKELTVKEDKGAQSGSGSATSGKSEATSGSSSSSAKSGTGEASGSSGASSSSG		846
Qy	1237	QSRSGOPKA-----ATCKVSSINQETIOTYCVEDTPICFSCSSLSLSLSAEDETG		1288
Db	847	SGVSGESSSVSTESGSGFTSGSGVGEATGTVGDSE-----SGKPSKSSSTEELKLP		900
Qy	1289	CNQTTQEAADANTLOTAIEIKERIGTRSAEDPVSEYPAVSHQHPRTKSSRLQSSLSSESAR		1348
Db	901	FTKNGEKSPIGS-----DITGKESSEETTSRKPIEGSDSLTEGSG-SGEWFTGSGK		952
Qy	1349	HKAVFESSGA-----KSPSKSGAQ-----TPKSP--PEHYVQETPLMESRCTSVS-		1391
Db	953	H-----FESGKSVTSVTKGTQSGAEGSGKPKPGGAPG-----ITTDGESSSTGTD		1005
Qy	1392	-----SLDSFESRSIAS-----SVOS		1407
Db	1006	KSGGPKADKDNKNVKTGDKNPDITTDGEDTSETSGGEGQPKGKSGQPPDGKGEVK		1065
Qy	1408	EPCS-----GMVSGI-----ISPDLDLP-----SPG-----QTMPPSR		1435
Db	1066	KPTSEVDGPNLSGTGKGNVLPKPTDLPESGILITSSGKNSTFEHGTKLRLPP--		1123
Qy	1436	SKTPTTPTTQATKRVKPKNKAAPTAE---KRSQPKQAANAQVQVLPADITLLHFA		1492
Db	1124	-KTEDKSSETPQGLGISAKKPEPEDGTSGEVG-----LEILWESTTSGSTT		1170
Qy	1493	TESTPDGFCSSLSALSLEPFIQ-----KQVELRIMPVQVQNDNGNE- 1536		
Db	1171	LDSDSVGLSIGSDLTAKTKKPHVEIEGSGTGDEETATTRDYSKSTKPRVEVDGDN		1230
Qy	1537	-----TESEQKESNENQEKAEKTIQSEK-----LLDSDDDDDIDILEECII		1580
Db	1231	ETSGVDGKPTTAPTPSSSAESTSRIPITSESPGSGGEGAGVPSPGSGESSTAPD		1290
Qy	1581	SAMPTKSSRKAKPAQATASKLPP-----PVARKPSQLFVYKILPSONRLQPKHVSFT		1633
Db	1291	GVSTTSATAPVPTTSASSTPDAVEESGIPSSKTAEPLETTAPSTEVTSPEG--SGT		1348
Qy	1634	PGDMPRVYCVGT-PINFSTATSLDLTIESPNELAAEGVR-----GGAQSG		1682
Db	1349	EESTLPP--TEGSGESTTSSAPTVEPATVLPQNRNEKPEPTKOTFALPTTTTGAPOAN		1405
Qy	1683	EPEKRDITPEGRSTDEA-----QGG--KTSSVTIPE-----		1712
Db	1406	DSSVENTKCT---SSDECGLDALCERRTGVCRCPEGEGAPPKSCVDVDECATGDHNC		1462
Qy	1713	-----LDDNKAEE-----GDILAECI		1728
Db	1463	ESARCONVGVYACFTGFRKADGSCQDIDECTEHNSTCCGANAKCVNKPQTYSCCE		1522
Qy	1729	NSAMPKGSKHKPFRVKIMDVQOASASSAPKN-QLD-----GKKKKPTSP		1775
Db	1523	NGFIGDGYQCVPTTKAPCDOSTGSKSHCSSESNVCEVDVDSVECKECMGYKKSQV		1581
Qy	1776	VKPIQNTYRTRVRKNADSKNINLAERFVSDNKKQNLKNS---KDFND-----K		1826
Db	1582	CEDINECVAEKAPCSLNA---NCVNMNGTFS---CSCQGYRGDGMCTDINECDEHPC		1635
Qy	1827	LPNNE-DRVGRGSPAFDSPHY-----TPTEGT-----PYCFS---RND		1860
Db	1636	HPHAECTNLEGSFKCECHSGFEGIGIKKTNPLERSCEDVEKFCGRVDHVSLSVRYNG		1695
Qy	1861	SLSSL-----DFDDDDVDLSREKAELEKA-----KENKESEA		1892
Db	1696	SLSVCECEPCGFRFEKBSNCVDIDECESRNCCDPASAVCVNTEGSRCECAEGEG		1755

Db 622 SASSTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASAST 681
 QY 1094 CQCCVCPYRGANGBETNRVGNHGINQNVSQLCOEDYEDDKFTNYSEYSEEQH 1153
 Db 682 SASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASEGASTSA-- 739
 QY 1154 EEEERPTNYSIKYNEKRHRVDQIDYSLKYATDIPSSQKQSFPSKSSGQSSKTEHMS- 1212
 Db 740 -----SASASTSASASASTSASASASTSASASASTSASASASTSASA 773
 QY 1213 -----SSSENTSTPSGNKRONQLHPSQAQBSGQFOKAATCKVSSINOETIYQVCDTP 1268
 Db 774 SASISASESASTSAS-----ASASTSASASASTSASASASTSASESASTSASASAS 824
 QY 1269 ICFRCSSLSLSAEDEIGCNQTTQADSANTLIQIAIEKIKIGRGAEDPVSEVPVSQ 1328
 Db 825 TSASASTSASASASTSASASASTSASASASTSASA-----SASTSASESASTSASASAS 880
 QY 1329 HPRTKSRLOQSSLSSEARHAKVEFGSAKPSKS-CAQTPKPPBHYVOETPLMFSCR 1387
 Db 881 TSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSAS 940
 QY 1388 TSVASLDSFEFRSTASSVQSEPCGMVGIITSPDLPSDQTPMPPSRKTPPPPPOTAQ 1447
 Db 941 ASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASAST--SASASAS 998
 QY 1448 TKREVPKNKAPTAKRESGPQAAVNAVQVLPDADTLHPATESTPDGFCSSLS 1507
 Db 999 TSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSAS 1050
 QY 1508 ALSLDEPIQKDVLRIMPPVQENDNGNETSEBQPKESNENQEKAKTIDSEKDLDDS 1567
 Db 1051 A-----SASTSASESASTSASASASTSASASASTSASESASTSASASASTSAS 1090
 QY 1568 DDDDIIELEECIISAMPTKSRKAKKPAQTAKLPPPVARKPSQLPVYKLLPSONRLQPQ 1627
 Db 1091 ASASISASESASTSASASASTSASASASTSASASTSA-----SSEASTST 1137
 QY 1628 KHVSTPGDDMPRYCYVEGTPINFSTATSLSDLTIESPPNELAAGEVGRGAQSGEPEKR 1687
 Db 1138 SASASTSASES-----ASTSASASASTSAS-----ASASTSASASASTSASASASTSASA 1187
 QY 1688 DTIPTEGSTDEAOGGTSSVTIPELDNKAEEGDILAECSNAMPKQKSHKPRVVKIM 1747
 Db 1188 STSASESASTSASASASTSASASASTSASASASTSASASASTSASASTS----- 1238
 QY 1748 DQVOQASASSAPNKNQLDGKKKTPVKPIQONTYRTRVRKNADSKNNLNARVFPD 1807
 Db 1239 ---ASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASA 1295
 QY 1808 NKDSKQNLKNNKDFNDKLPNNEDRVGSAFADSPHHYIEGTPYCFSRNDSLSSLDP 1867
 Db 1296 SASASTSASVSASTSASESASTSAS-ASASTSASASASTSASESASTSASESASTSASES 1354
 QY 1868 DDDVDLAREKAEKAKENKESAKVTSHTELTNSQOANKTQIAIKQPINRGOPKPI 1927
 Db 1355 ASTSASASASTSASASTSASASTSASASTSASASTSASASTSASA----- 1401
 QY 1928 QKQSTFPQSKDIPRGAATDEKLNFAIENTPVCFSHNSLSLSDIDQENNNKENPPI 1987
 Db 1402 SASTSASASASTSASASASTSASASTSASASTSASASTSASASTSASASTSASASASI 1461
 QY 1988 KETEPDPSQGEPS-KFQASGYAPKPFHYE-DTPVCFVRNSSL-SLSLDSDDLQECIS 2044
 Db 1462 SASESASTSASASASTSASASTSASASTSASVSASTSASASTSASASASTSASESASTSASA 1521
 QY 2045 SAMPKXKXPSRLKGNKHSPPRMNGGILGEDITLDKDIQRPDSEHGLSPDSENFDMKAI 2104
 Db 1522 SASTSASASASTSASASASTSASASASTSASESASTSASESASTSASASTSASESAST 1581
 QY 2105 QGANSIVSLHQAACLAACLSQAQSDSDSIILSKSGTSLGSPFLTPDQEKPTSKNG 2164
 Db 1582 SASASTSASASTSASASTSASASTSASASTSASASTSASASTSASESASTSASESASTSASA 1641

QY 2165 PRILPGEKSTLETTKIESESGIKGGKVKYKSLIT---GKVRNSNSEISGOMKQPLQANM 2221
 Db 1642 SASTSASASTSASASASTSASASTSASVSASTSASESASTSASASTSASASTSASAST 1701
 QY 2222 PSISGRTHIHPGVNRSSSTSPSVKGPPLKHTPASKSPSEGQATTTSPRGAKPSVKSE 2281
 Db 1702 SASESASTSASASTSASASTSASESASTSASESASTSASASTSASASTSASASTSASA 1761
 QY 2282 LSPVAREQTSGIGSSKAPSRSGRSTPSRPAQOPLSRPIQSPGRNSISPGRNGISPPNK 2341
 Db 1762 SASASTSASASTSASASTSASASTSASASISASESASTSASESASTS-----ASASTS 1816
 QY 2342 LSQLPRTSPSTASTKSSGSGKVSYTSPQRMSQONLTQOTGLSKNNAIPSEBSAKGL 2401
 Db 1817 ABEASTSASASTSASASTSASA-----SASTSASASTSASESASTSASESASTSAS 1872
 QY 2402 NOMNNGANGANKVELSRMSSTKSSGSESDRSERPVLRQSTFIKAPSPTLARKLEESAS 2461
 Db 1873 ASASTSASASTSASASTSASASTSASASTSASVSASTSASASTSASASTSASESASTS 1932
 QY 2462 FESLSPSSRPASPTRSQAOQTPVLSPLPDMSLSTHSSVQAGGWRKLPPLNLSPTIENDGR 2521
 Db 1933 TSASASTSASASTSASASTSASASTS-ASASASTSASA----- 1971
 QY 2522 PAKRHDIAKSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSSILSASES 2581
 Db 1972 -----STSASESASTSASASTSASASTSASASTSASASTSASASTSASASTSASA 2024
 QY 2582 EKAKSEDEKHNVISGTKQKQENQVSAKGTWRKIKENBPSPTNSQTIVSSGATNGAESK 2641
 Db 2025 ESASTSASASTSASASTSASASTSASASTSASESASTSASASTSASASTSASASTSASA 2084
 QY 2642 TLIYQMAPAVSKTEDVVMRIEDCPINNPRSGRSPONTFPVIDSVSEKANPNIKOSKDQ 2701
 Db 2085 T-----SASASTS-----ASASTSASASTSASESASTSASESASTSASASTS 2126
 QY 2702 AKQNVGNSVPMTVGLNRLNSFIQVDAPDQKGTETIKQNNPVFVSTNESSIVERTP 2761
 Db 2127 ASESA-----STSASASTSASASTSASASTSASASTSASESASTSASASTS 2166
 QY 2762 FSSSSSKHSSPSGTVAARVTPFNYPNPSRKSADSTARPQIPTFVNNTKKRDSKTD 2821
 Db 2167 ASASTSASASTSASASTSASESASTSASESASTSASESASTSASESASTSASESASTSASA 2220
 QY 2822 STSSGTSQSPKRHSGS 2837
 Db 2221 ASASTSASASTS 2236

RESULT 13

T02345
 Hypothetical protein KIAA0324 - human (fragment)

C:Species: Homo sapiens (man)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C:Accession: T02345
 R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 submitted to the EMBL Data Library, March 1998
 A:Description: Sequencing of human chromosome 16p13.3.
 A:Reference number: Z14664
 A:Accession: T02345
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1791 <R1C>
 A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AC08453.1; PID:g2996650
 C:Genetics:
 A:Map position: 16
 A:Introns: 1610/2; 1706/2
 A>Note: KIAA0324

Query Match 2.9%; Score 426.5; DB 2; Length 1791;
 Best Local Similarity 20.0%; Pred. No. 9.4e-10;

Matches 382; Conservative 251; Mismatches 744; Indels 533; Gaps 78;
 QY 1210 HMSNSENNTS-----TPSSNAKQNLHPSSAOSRSQPOKAATCKVSSI 1254
 Db 1 HSGSSSDTKVKPTPRQSHSGSISYPKVKQAQTPPGPSLSGSKPCP-----49
 QY 1355 NQETIQYCVEDTPICFRSCSL-SLSSABDEIGN-----QTQEAADSANTLQIAEI 1307
 Db 50 -QEKSDSLVQSCPSGLSLCAGVKSTPPGESYFVGSLLQKQSQSTSPDHRSDTSSPEV 108
 QY 1308 KEGIGTSAEDPVEPAVSHQHPET--KSSRLQSSLSSESARHKAVERFSSGAKSPS---1362
 Db 109 RQS-----HSEFSLQSKQSTPKGGRSRSSSPVTELASRPIRQDRGFASPMPL 159
 QY 1363 KSGAOTPKSPPEHYVQTPLMFRCSTSVSLDFEGRSIASSVQSPFCGMSGIIISP-S 1421
 Db 160 KSG-----MSPEQSRFQ-----SDSSSYPTVDSNLSLQGRLETAESKERWALPPQE 206
 QY 1422 DLDPSCQTPPPRSKTPPPPPQT-----AQTREVPKNAKPTAEKRESGPKQAAVN 1473
 Db 207 DATASP-----PRQDKSPFPVQDRPSSLVFQDTPRERSGAGSGSPETKQNSALP 262
 QY 1474 AAVQRVQL-----PDADTLHFAT-----ESTPDGFCSSSL-----ALSDEPFIQ 1517
 Db 263 TSSQDEELMEWEKSEBPAGQILSHLSSELUENKSTNFFESSPEVERPAVSLTLDQSQS 322
 QY 1518 KDVELRMP-----PVQENDNGN-----ETESQPKESNENQKEAB 1554
 Db 323 ASLEAVEVPMASWGGPHFSPBEKLSNPLRENSFGSPLEFRNSGPGTGEMNTGFSSE 382
 QY 1555 KTDSEKDLDDSDDDDIETLECIISAMPTKSRKAKPAQATASKLPPPVAVKPKQLPV 1614
 Db 383 VKEDLNGPFLNLETPSLDMKE-----QSTRS-----GHSSSELSDAVEKAG-----427
 QY 1615 KYLLPSQNRLOPKHVSFTPGDMPRYCVGPTPIINFSTATS--LSDLTIESPPNELAAG 1672
 Db 428 -----MSSNQSISSPVLDAVPTPSRERS-----SSASPEMKDGLPRTPSRRSRG 474
 QY 1673 E--GVRGAQSGBEFKRDTIPTEGRTDQAQGTSSVTIPELDNKAEEGILAECSN 1730
 Db 475 SFGLDGGS-----GTSRHLSGSS-----PGMKDIPRTPSRGRSECDSS 515
 QY 1731 AMPGKSHKFRVKINDVQQAASASSAPNKNQ--LDGKKKKPTSPVKPIQNTYEYTR 1788
 Db 516 PEPKALPQT-----RPRSRSFSPPELNNKCLTPQERSGSE--SSVQKTVAETP 564
 QY 1789 V--RKNADSKNNAERVDNDKSKQNLKNSKDFNDKLNEDRVGSPAFSPHY 1846
 Db 565 LQGRSRGSGQELDVKPSAPQERSE-----SDSSPDSKAKTR-----602
 QY 1847 TPIETGYCFSRNDSLSSLDFFDDVLDGREKAEKAKENKESAKVTSTELTASNOQS 1906
 Db 603 TPLR-----QRSRSGSPEDV-----SKRSLSPRSRSGSPPEVKDKPRAAPRAQS 648
 QY 1907 ANKQIAKQPINRGOPKILQKSTFPOSSK--DIPRGATDEKLNFAIENTPVCF 1963
 Db 649 GDSDSPEPKAPAPRALP-----RRSRSGSSKGRGSPGEGSSTES-----SPEHPP---695
 QY 1964 SHNSSLSLSLDIQENNNKENEP-IKETEPDPSQCEPKPQASGVAPKSFHYED-----2016
 Db 696 -----KSRARGRSRSPKTKSRTPPFRSRSSSPPELTRKARLSRRSRSSASSPE 747
 QY 2017 -----TPVCFSRNSSLSDIDEDDLQECISAMPKKKKPRLKGDNEKHSPRNMGIL 2072
 Db 748 TRSRTPPFRHRSFVSPEPAEKSRSSRRRRGASSPRTKTSR-RGRSPSPKPRGLQSR 806
 QY 2073 GEDLTLDLQIDPDEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLSRQASDS 2132
 Db 807 SRSRREKTTTRRD-----RSGSSQSTR-----RRRGRSR 839
 QY 2133 DSILSKSGISGSPPH-LTPQEEKFTPSNKGPRILKPGEXSTLETKKIESEKIRGG 2191
 Db 840 SRVTRRRRG---GSGYHSRSPARQESSRSTSSRRR---GRSRTPTTKRSRSTSPAP 892

QY 2192 KKVYKSLITKVRNSEISGQMKOPLOANMPSISGRRTMIHPCGVRNSSSSTSPVSKKGP 2251
 Db 893 WK-----RRSRASPATHRRSRRTPLISR-----RRSRSTSPVSRRS 932
 QY 2252 PLKTPAKSPBEGQVATTSPRCAKPSVKSELSPVARTQTSQIGSSKAP--SRSGSRDSTP- 2309
 Db 933 RSRTSVTRRRSRASPVSRRS---RSRTPPVTRRRSR---SRTPTRRRSRSTPP 984
 QY 2310 ---SRPAQOPLSRPIOSPCGNSISP-----GRNGISPPNKLQPLRTSPSPSTASYKSS 2359
 Db 985 VTRRRSRRTPPVTR---RRSRSTPITRRSRSTSPVTRRRSRSTPVTTRRRSR 1041
 QY 2360 GSGKMSYTPSGRQMSQQNLT-----KOTGLSKQASSIPRSEASKGLNWNNGA 2410
 Db 1042 ---TSVTRRRSRRTPPAIRRRSRRTPLPKRSRSRSLAIRRRSRSTPRTA 1094
 QY 2411 NKKVELSR---MSSTKSSGESDRSR-----PVLV---RQSTFIKEAP 2448
 Db 1095 RGKSLTRSPAIRRRSAGSSDRSRSATPPATRNHSGRTPPVALNSGRMCSFRPSM 1154
 QY 2449 SPTLRRKLEESAFESLSPSRSPASPTRSQATPVL---SPSLPDMSLST---HSSVOA 2501
 Db 1155 SPTPLDRCSRPGMLEPLGSSRTPM--VLQAGGSMMDGPGPRPDHQTSPVENHQSRI 1213
 QY 2502 G-----GWRKLPNLSPTIENDGRPAKHDIARSHSESERLPIKNSGTWKRH-- 2551
 Db 1214 ALALTAISLGTARPPMSAA-----GLAARMQVPAPVPLMSLRTAPAANLA 1261
 QY 2552 ---SKHSSSLPRVSTW---RRTGSSSILSASSSESEKAKSEDEKHVN-- 2593
 Db 1262 SRIPAASAAANLASARTPAIPTAVNLADSRTPAAAAMNLAASPTAVAFSA---VWLA 1317
 QY 2594 -----SISGTOSKENQVSAKGT--WRKIKENFSP--TN 2624
 Db 1318 DPRTPTAPAVNLAGARTPAALASLGTSGTPTTAANYPSSRTPQAPASANLVGPSAH 1377
 QY 2625 STQTSVSGATNG--AESKTLIVQMAPAVKTEDVWVRI-----2661
 Db 1378 ATAPNLAGRTAAALPASLTSARMAPALSGANLTSRVLPSAYERVSGRTSPPLDRA 1437
 QY 2662 -----EDCPINPRSGRSTGN-----TPPVIDSVSEKANPNKD 2696
 Db 1438 RSRTTPSAPSQSRMTSERAPSPSSRMGQAPSQSLPPAQDQPRSPVPSAFSDQSRCLIAQ 1497
 QY 2697 SKDQAKQNVGNGSVPRVTYGLENLNSFIOVDAPDKGTETI-----KPCQNP- 2745
 Db 1498 TTPVAGSQSLSSGAVATTTSAGDH--NGMLSVPAPGVPHSDVGEPPASTGAQOQPSALAAL 1556
 QY 2746 VPVSETNESSIVERTPPSSSSSKHSSPSGT-----VAARVTPPENYN 2787
 Db 1557 QPAKERSS 1613
 QY 2788 PSPKSSADSTSRAPSOIPTPVNNNTKRDSTKTDSTESSGTQSPKHHGS 2837
 Db 1614 PTP---APKEAVREGRPPEPTPAKRRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1659

RESULT 14

T30249
 cell proliferation antigen Ki-67 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30249
 R:Starborg, M.; Gell, K.; Brundell, E.; Hoog, C.
 J. Cell Sci. 109, 143-153, 1996
 A:Title: The murine Ki-67, cell proliferation antigen accumulates in the nucleolar and het
 1 for cell cycle progression.
 A:Reference number: 220787; MUID:96431717; PMID:8834799
 A:Accession: T30249
 A>Status: preliminary; translated from GS/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-2938 <STA>

A:Cross-references: EMBL:X82786; NID:g1177527; PIDN:CAA58026.1; PID:g1177528
A:Experimental source: strain CBA; testis
C:Genetics: Ki-67
C:Keywords: cell cycle control; nucleus; tandem repeat

Query Match 2.9%; Score 420; DB 2; Length 2938;
Best Local Similarity 18.6%; Pred. No. 3.3e-09;
Matches 604; Conservative 439; Mismatches 1268; Indels 932; Gaps 147;
QY 80 SNFPVGLRSKXVSRSYSGREGSVSSRGECSPVMGSGF--PRRGFV-----124
DB 109 TEFFPKSLGKEPSRA--SRDSFCADPDGEGQDTKASKWTASRRSFVYAKGLSADSPASD 166
QY 125 ---NGSRESGYLEE-----LEKERSILLADLDKEEKEKD 156
DB 167 GSKNSVSDSGSHVEQHTGRNIVEPTSGSLRSPGLQCAVGTGRSLLPOTOSLSNSKE 226
QY 157 WYIAQL-QNLTKRID-----SLPTENFSLQTDNTRRLQLEYEAR-----QIRVA 199
DB 227 SPFEKLYQSMKEELDVKSQKCRKSEPOPDRAABESRETQLLVSGRAFAKSGGTPVTAA 286
QY 200 MEEQLG-----TCQMEKRAQR--IARIQIEKDIILRILOLOSQA TEAERS 245
DB 287 SSPKVGKIWRWRGMPVPTSTETAKMKTFRHSQQLKDEDSRV-----TGRRH 337
QY 246 SONKHETGSHDAERQEQGVCEINAT-----SGNGQGSTTRMDHETASVLSGSSSTH 298
DB 338 SVNLDGSGSAQVHKTVTPG---KLATFNQTPVEAGDVGSPAUTPEH-----SSSPQR 387
QY 299 SAPRL--TSHLGKVMYVSLMSLGLTHDKDMSRLLTLLAMSSQSCISMROGCLPL 355
DB 388 SIPAKVEAPSAETQNLRLTQLRFGKEKTKPGSFKPEKLATAAEQTC-----SG-LPG 441
QY 356 LIQLLGHNDKDSVLLGNRSGKEARARASAAHL-----NIHDSQDDKRGRRRIARVL 407
DB 442 LSSVDISNFGSI--NKSEGMKRRRVSGHRLRPELFDENLPNTP-LKRGTPTKRK 498
QY 408 HLEQIRAYCTEWQBAHPGMDQDNMPAPVEHQICPAVCVLMKLSFDEHRHAMN 467
DB 499 SLGTHSPAVLTIK-ERPQSPGQESPGITPRTN-----DQRRRGST 542
QY 468 ELGGLQATAEILLQVDCEMYGLTNDHYSITLRYAGMALNTLTFDGVANKATLC-SMGCM 526
DB 543 SSG-----SNLCETDIPKAGKSGNLP-----AKRASISRSQHIL 580
QY 527 RALVAQLKSESDLOQVIASVLRNLSPADV---NSKKTLEVGSVKALMECALVEVKES 583
DB 581 QMICKRSRSGASEANLIVAK-----SW-ADVVLGVKQTKVAKHVPPKQTSKRQRPS 634
QY 584 TLKSVLSALWN-----LSAHC-----ENKADICAVDG 611
DB 635 TPKEFTSLNHNQFTTGHSNPFCTIVVGRAQIEKVSVPARPYKMLNMLNKRKVDSEDS 694
QY 612 ALAFVLGLTVRSQNTLAIETESGGILRNYSLL-----IATNED-----651
DB 695 GLTEFKTPVKEKQQ-----MSDTGVSLSNANLSERQLVNTNSGDIPETITILGEKV 750
QY 652 -----HROILRENNCLQTLQHLKSHSITVSNACGTLWNLSAR 690
DB 751 LSSTRNAAKQSDRYASAGFTLRRRSIKHENTVQT---PKNVHNIT-----DLEKK 797
QY 691 NP-KDOEALWDMGAVSMKLNLIHSKHKM-----AMGSAALRLMA-----731
DB 798 TPVSETEPLKTVASSVKLRRSRELRHLIVETVNEKTEAVLAENTTAHRLGTFTEQKVQD 857
QY 732 -----NRPKYKDANIMSPGSLPSLH---VRKQKALEAELDAQLHSETPDNIDNLSPK 782
DB 858 QVQDNENAPQCKESGELSEGSEKTSARRSSARKQKPTKDLGSGQWVQTADYAEALLSQ 917
QY 783 ASHRSQKHQSLGYDVFDRNHDNDNSDNFNGNMTVLSPLYLNTVLPSSSSSRGSGLD 842
DB 918 G-----QGTIQNL-----BESMMQNTSISGEOG---ITEKKVNIIVYATKHXSPKTP 963

QY 843 SRSE-----KORSLERER-CIGLGNVHPATENPGTSSKRQLQISTTA 884
DB 964 GKIAQLEAGLKEHFEFETPNPKDKBITEDRVLCKSPQVTITTTKP--QTSISG 1021
QY 885 AQIAKVMEEVSA---IHTSQEDRSSGSTITLHCVTDERNALRRS-----SAATHSNT 934
DB 1022 KKV-DMEKSSALTKEIHPGESRHPKILKEC--EDIKALKQSENEMLTSTVNGSKRT 1078
QY 935 YNFTKSEN---SNRTC-----SMP---YAKLEYKRSSNDLSNSVSSDGYGKRGOMKP 981
DB 1079 LGSKKKAQPLEDLTCQELIFISPVPTNIKKIPSKSPHTQPVRTASTKRLSKTGLSKV 1138
QY 982 SIESYSEDDSKCSYGOVPADLAHKHS-----ANHMDDND--GELDTPIVSLKY 1031
DB 1139 DVR-----QEPSTLGRKTKSPGRAPGTPAPVQBEENDCTATMETP-----1177
QY 1032 SDEQLNGRQSPQNERWARPKHI--IEDEIKOSEQORQSRNOSTTYPVYTESTDDXHLKF 1089
DB 1178 -KQLESIENTLGLRKQSRTPKXDTITGFQDSFOIPDHANG-----PLVVVKT--KKMFF 1227
QY 1090 QPHFGQOECVSPYRSGANGSETRNVSGNHGNGVNSQSLCOEDDYEDDDKPTNYSRYSE 1149
DB 1228 --NSPOPESAITKSR-----EROSRASISKIDVKE-----ELLES 1261
QY 1150 ECHHEEERPTNYSIKYNEBEKRVHDQPIDYSLKYATDIPSSQKQSFQSKSSGQSKTE 1209
DB 1262 EEHLQLEGVDTFQVSTNKVIRSSRKPAKRLDSTAGMPNSKEMRC-----SSKDNTPCLE 1317
QY 1210 HMSSESENTSTP-----SSNAKQNLHPSSAQSRSQGPQKAATCKVSSINQET 1258
DB 1318 DLNGFOELQMPGYANDSLTTGISTMLARSPQLGPVRTQ-----INXKS 1361
QY 1259 IQYCVYEDTPICFSCSSLSISSAEDIEGCTOTQEADSANTLQIAEI-KEKI-----G 1312
DB 1362 LPKLILKMDV---TEBISGL--WKQSLGRVHTTQEQEDNAIKAMEIPKETLQTAADG 1415
QY 1313 TRSAEDPVSEVPVAVSQHPRTKSRLOGSSLSSESARHKAVERFSSGAKSPSKGAQTPKGP 1372
DB 1416 TRLTROP-----QTPKEKQVPLEDHSHVFOE--LFQTSRYCSDPLIGNKQTRMSLRSP 1465
QY 1373 PEHVQETPLMFRCSTSVSLSDFESRSIASSVQSEPC--SGMVSGI-ISPDS-----LP 1424
DB 1466 QPFVFR-TPTSRLAK--TSVGNIAVREKISPVSLPQCAGTEVVHIPIGEDDTEKNGVK 1523
QY 1425 DSPGQMPSPSRKSTPPPPQTAQTKREVPKNKAPTAEKRE-----SGPKQAANAAVOR 1478
DB 1524 ESTPQILDSSASRTVSKRQQAHEER--PQSGDLPHQELFQTPASGKDPVTVDSTTKI 1581
QY 1479 VQVLPDADTLILHPATESTPDGSCSSLSALSILDEFFIOKDVLR-----IMPPVQENDN 1533
DB 1582 ALQSPQPGHIINPASVKRQSNMSLRKDMREFSI---LEKQTSQSGRDAGTPAPMOE-EN 1636
QY 1534 GNETESEQPK-----SNEQKEAEKTTIDSEKDLLDDSD-----DDDEILLE 1576
DB 1637 GTTAIMETPKQLDFFGNSTGHKRRPRTPKNRAQPLEDLDGQELFQTPAGADPVSVEE 1596
QY 1577 ECTIS-----AMPTKSRKAKPAQTA-SKLPPPVARKPSQLPVYKLLPSQNLQPKHV 1630
DB 1697 SAKISLASSQAEPVPTASTKRRSKTGLSKV--DVRQEPSTL-----1736
QY 1631 SFTPGDMPRVYVEGTPI-----NPFSTATSLDLTIESPPNELAAGEVGRGQAQGEFE 1685
DB 1737 ---GRKMSLGRAPGTAPVQBEENDSTA-----FMETPKQKLD-----TGNSSGHR 1781
QY 1686 KRDTIPTGRSTDEAQ-----GKTSSTVPIPE-----LDDNKAEGDILAEICNSA 1731
DB 1782 RPQTPKIRAQPLEDLGQFQELFQTPAGANDSVTVBESVKMSLESSQAE--PVKTPASTKR 1839
QY 1732 MPK-GKSHRPFRVKKIMDQVQOASASSAP-----NKMQLD-----GKKKK 1771
DB 1840 LSKTGLSKVDVREDPSILEKTKSPGTAPVQBEENDCTAFMETPKOKLDFTGNSSGHKR 1899

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:21 ; Search time 32.5 Seconds
(without alignments)

4554.937 Million cell updates/sec

Title: US-09-442-489F-7

Perfect score: 14566

Sequence: 1 MARASYQLLKQVEALKMEN.....ESSGTQSPKRHSGLVTSV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14566	100.0	2843	1 APC_HUMAN	P25054 homo sapien
2	13198.5	90.6	2842	1 APC_RAT	P70478 rattus norv
3	13178	90.5	2845	1 APC_MOUSE	Q61315 mus musculus
4	476	3.3	2774	1 MAPA_RAT	P34926 rattus norv
5	474	3.3	5560	1 SPEN_DROME	Q8sx83 drosophila
6	452	3.1	3664	1 MINT_HUMAN	Q96t58 homo sapien
7	428.5	2.9	3644	1 MINT_MOUSE	Q42504 mus musculus
8	404	2.8	2688	1 ZEP1_MOUSE	Q31172 mus musculus
9	401.5	2.8	2805	1 MAPA_HUMAN	P78559 homo sapien
10	399	2.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
11	391.5	2.7	3562	1 PCV1_CHICK	Q90953 gallus gall
12	390	2.7	5085	1 PCLO_RAT	Q9jks6 rattus norv
13	389.5	2.7	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
14	380.5	2.6	4911	1 MLL3_HUMAN	Q8ne24 homo sapien
15	379.5	2.6	3210	1 CENF_HUMAN	P49454 homo sapien
16	375.5	2.6	2459	1 MAPB_RAT	P15205 rattus norv
17	374.5	2.6	4377	1 ANK3_HUMAN	Q12955 homo sapien
18	374	2.6	3256	1 KI67_HUMAN	P46013 homo sapien
19	372	2.6	5120	1 PCLO_CHICK	Q9pu36 gallus gall
20	371.5	2.6	1581	1 PRB_HUMAN	Q15648 h peroxisom
21	370.5	2.5	3969	1 HFX_HUMAN	Q03164 homo sapien
22	370	2.5	2453	1 NCR1_MOUSE	Q60974 mus musculus
23	365	2.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
24	365	2.5	2440	1 NCR1_HUMAN	O75376 homo sapien
25	363	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
26	363	2.5	3259	1 GCB1_HUMAN	Q14789 homo sapien
27	362.5	2.5	2663	1 CENF_HUMAN	Q02224 homo sapien
28	362	2.5	2492	1 ATRX_PANTR	Q7yqmd pan troglod
29	361	2.5	7389	1 PFAL_MOUSE	Q9izuc mus musculus
30	359	2.5	5038	1 PCLO_MOUSE	Q9gyx7 mus musculus
31	358.5	2.5	6632	1 UNH9_CAEEL	O01761 caenorhabdi
32	357	2.5	2464	1 MAPB_MOUSE	P14873 mus musculus
33	356.5	2.4	1972	1 P531_HUMAN	Q12888 homo sapien

ALIGNMENTS

RESULT 1

ID	APC_HUMAN	STANDARD	PRT	2843 AA
AC	P25054; Q15162; Q15163;			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Adenomatous polyposis coli protein (APC protein).			
GN	APC OR DP2.5 (Human)			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91335210; PubMed=1651562;			
RA	Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,			
RA	Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,			
RA	Finnear R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,			
RA	Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,			
RT	"Identification of FAP locus genes from chromosome 5q21.";			
RL	Science 253:661-665(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=91330307; PubMed=1678319;			
RA	Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,			
RA	Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,			
RA	Sargeant L., Kravcho K., Wolff E., Burt R., Hughes J.P.,			
RA	Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,			
RA	Abderahim H., Cohen D., Leppert M., White R.;			
RT	"Identification of deletion mutations and three new genes at the			
RL	familial polyposis locus.";			
RN	[3]			
RP	ASSOCIATION WITH CATENINS.			
RX	MEDLINE=94082295; PubMed=8259519;			
RA	Su L.-K., Vogelstein B., Kinzler K.W.;			
RT	"Association of the APC tumor suppressor protein with catenins.";			
RL	Science 262:1734-1737(1993).			
RN	[4]			
RP	DISEASE.			
RX	MEDLINE=95174843; PubMed=7661930;			
RA	Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,			
RA	Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,			
RA	Wood P.A., Taqi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,			
RA	Termette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;			
RT	"The molecular basis of Turcot's syndrome.";			
RN	[5]			
RP	DISEASE.			
RX	MEDLINE=97094176; PubMed=8940264;			
RA	Eccles D.M., van der Luijt R.B., Breukel C., Bullman H., Bunyan D.,			
RA	Fisher A., Barber J., du Boulay C., Primrose J., Fodde R.;			
RT	"Hereditary desmoid disease due to a frameshift mutation at codon 1924			
RT	of the APC gene.";			

P08640 saccharomyc
P55200 mus musculus
P30414 homo sapien
Q03661 saccharomyc
Q9v618 h nuclear r
P30415 mus musculus
P46100 homo sapien
Q9wu42 mus musculus
Q95923 rattus norv
Q9nzw4 homo sapien
Q7yqmd3 pongo pygma
P47035 saccharomyc

- Am. J. Hum. Genet. 59:1193-1201(1996).
[6]
RN DISEASE.
RP MEDLINE=20243021; PubMed=10782927;
RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
RA Rodde R., Almon B., Bapat B.;
RT "A germline mutation at the extreme 3-prime end of the APC gene
RT results in a severe desmoid phenotype and is associated with
RT overexpression of beta-catenin in the desmoid tumor.";
RL Clin. Genet. 57:205-212(2000).
[7]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
RP MEDLINE=20384842; PubMed=10926498;
RA Day C.L., Alber T.;
RT "Crystal structure of the amino-terminal coiled-coil domain of the
RT APC tumor suppressor.";
RL J. Mol. Biol. 301:147-156(2000).
[8]
RN X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
RP CTNBB1.
RX MEDLINE=21564054; PubMed=11707392;
RA Ekiof Spink K., Fridman S.G., Weiss W.I.;
RT "Molecular mechanisms of beta-catenin recognition by adenomatous
RT polyposis coli revealed by the structure of an APC-beta-catenin
RT complex.";
RL EMBO J. 20:6203-6212(2001).
[9]
RN X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
RP AXIN.
RX MEDLINE=20271867; PubMed=10811618;
RA Spink K.E., Polakis P., Weiss W.I.;
RT "Structural basis of the axin-adenomatous polyposis coli
RT interaction.";
RL EMBO J. 19:2270-2279(2000).
[10]
RN REVIEW ON VARIANTS.
RP MEDLINE=94154728; PubMed=8111410;
RA Nagase H., Nakamura Y.;
RT "Mutations of the APC (adenomatous polyposis coli) gene.";
RL Hum. Mutat. 2:425-434(1993).
[11]
RN VARIANTS FAP.
RX MEDLINE=91335211; PubMed=1651563;
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
RT patients.";
RL Science 253:665-669(1991).
[12]
RN VARIANTS FAP.
RX MEDLINE=93265030; PubMed=1338904;
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
RA Miki Y., Mori T., Nakamura Y.;
RT "Somatic mutations of the APC gene in colorectal tumors: mutation
RT cluster region in the APC gene.";
RL Hum. Mol. Genet. 1:229-233(1992).
[13]
RN VARIANTS FAP.
RX MEDLINE=93244793; PubMed=1338691;
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
RA Nakamura Y., Horii A.;
RT "Somatic mutation of the APC gene in gastric cancer: frequent
RT mutations in very well differentiated adenocarcinoma and signet-ring
RT cell carcinoma.";
RL Hum. Mol. Genet. 1:559-563(1992).
[14]
RN VARIANTS FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
RP SER-2502.
RX MEDLINE=93250848; PubMed=1338764;
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
Baba S., Nakamura Y.;
RT "Screening for germ-line mutations in familial adenomatous polyposis
RT patients: 61 new patients and a summary of 150 unrelated patients.";
RL Hum. Mutat. 1:467-473(1992).
[15]
RN VARIANT FAP TRP-99.
RP TISSUE=Peripheral blood lymphocytes;
RX MEDLINE=95134544; PubMed=7831149;
RA Dobble Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
RA Mueller A., Mueller H., Scott R.J.;
RT "Mutational analysis of the first 14 exons of the adenomatous
RT polyposis coli (APC) gene.";
RL Eur. J. Cancer 30A:1709-1713(1994).
[16]
RN VARIANT FAP GLY-722.
RP MEDLINE=95135430; PubMed=7833931;
RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Guanti G.;
RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
RT FAP patients.";
RL Hum. Mol. Genet. 3:1687-1688(1994).
[17]
RN ERRATUM.
RP Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Guanti G.;
RL Hum. Mol. Genet. 3:1918-1918(1994).
[18]
RN VARIANT FAP ILE-171.
RX MEDLINE=97144176; PubMed=8990002;
RA van der Luijt R.B., Meera Khan F., Vassen H.F.A., Tops C.M.J.,
RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
RA Plug R.J., Griffioen G., Fodde R.;
RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
RT familial adenomatous polyposis: 67 germline mutations identified by
RT DGGE, PTT, and southern analysis.";
RL Hum. Mutat. 9:7-16(1997).
[19]
RN VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
RX MEDLINE=98080146; PubMed=9419979;
RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
RA Fukutome A., Tomiyama J., Chuganji Y., Monoi M., Utsunomiya J.;
RT "Drastic genetic instability of tumors and normal tissues in Turcot
RT syndrome.";
RL Oncogene 15:2877-2881(1997).
[20]
RN VARIANT LYS-1307.
RX MEDLINE=98400248; PubMed=9731522;
RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
RT "The APC I1307K allele and breast cancer risk.";
RL Nat. Genet. 20:13-14(1998).
[21]
RN VARIANTS LYS-1307 AND GLN-1317.
RP TISSUE=Peripheral blood;
RX MEDLINE=98393712; PubMed=9724771;
RA Frayling I.M., Beck N.B., Ilyas M., Dove-Edwin I., Goodman P.,
RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
RT "The APC variants I1307K and E1317Q are associated with colorectal
RT tumors, but not always with a family history.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
[22]
RN VARIANT LYS-1307.
RX MEDLINE=98400259; PubMed=9731533;
RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
RT "The APC I1307K allele and cancer risk in a community-based study of
RT Ashkenazi Jews.";
RL Nat. Genet. 20:62-65(1998).

Query Match 100.0%; Score 14566; DB 1; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNMKEVLKQLQGSIEDAM 60
 Db 1 MAAAYDQLLKQVEALKMENSRLRQLEDNSNHLTKLETEASNMKEVLKQLQGSIEDAM 60

QY 61 ASSGOIDLLERLKEINLLOSNPPGVKLRKMSLRYSRGESVSSRGCECPVPMGSPPR 120
 Db 61 ASSGOIDLLERLKEINLLOSNPPGVKLRKMSLRYSRGESVSSRGCECPVPMGSPPR 120

QY 121 RGTVNGSRSTGYLBELEKERSLLMLADLDKEKEKDYWYAOQLNTKIDSLPTENFSL 180
 Db 121 RGTVNGSRSTGYLBELEKERSLLMLADLDKEKEKDYWYAOQLNTKIDSLPTENFSL 180

QY 181 QTDWTRQLEYEARQIRVAMEEQLQTCQDMKRAQRRIARIQOIEKDLIRQLQSOAT 240
 Db 181 QTDWTRQLEYEARQIRVAMEEQLQTCQDMKRAQRRIARIQOIEKDLIRQLQSOAT 240

QY 241 EAERSQNKHETGSHDAERQEGQGVGINNATSGNGGSTRMDHETASVLSSTHSA 300
 Db 241 EAERSQNKHETGSHDAERQEGQGVGINNATSGNGGSTRMDHETASVLSSTHSA 300

QY 301 PRRLTSHLGTKEVWYVLLSMLGTHDXDDMERTLLAMSSQDSCTSMRQSGCLPLLQLL 360
 Db 301 PRRLTSHLGTKEVWYVLLSMLGTHDXDDMERTLLAMSSQDSCTSMRQSGCLPLLQLL 360

QY 361 HGNDKDSVLLGNSRGSKEARASAAAHNIHSQDDDKRGREIRVLHLLQIRAYCETC 420
 Db 361 HGNDKDSVLLGNSRGSKEARASAAAHNIHSQDDDKRGREIRVLHLLQIRAYCETC 420

QY 421 WEWQBAEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGLQIAELLQ 480
 Db 421 WEWQBAEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEBHRHAMNELGLQIAELLQ 480

QY 481 VDCMYGHTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEEDL 540
 Db 481 VDCMYGHTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCMRALVAQLKSEEDL 540

QY 541 QQVIASVLRLNLSWRADVNSKKTLEVGSVKALMECALREVKESTLKSVLKALWNLASHCT 600
 Db 541 QQVIASVLRLNLSWRADVNSKKTLEVGSVKALMECALREVKESTLKSVLKALWNLASHCT 600

QY 601 ENKADICAVDQALAPLVGTLTYRSQNTLAIIESGGILNRVSSLIATNEDHRQILRENN 660
 Db 601 ENKADICAVDQALAPLVGTLTYRSQNTLAIIESGGILNRVSSLIATNEDHRQILRENN 660

QY 661 CLQTLLOHLKSHSLTIVSNACGTLWNLARSNPQOEALWDMGAVSMLKNLIHSHKHMIA 720
 Db 661 CLQTLLOHLKSHSLTIVSNACGTLWNLARSNPQOEALWDMGAVSMLKNLIHSHKHMIA 720

QY 721 GSAALRLNLMANRPAPKXKXANIMSPGSLPSLHVRKQKALELDAQHLSETFDNIDNLS 780
 Db 721 GSAALRLNLMANRPAPKXKXANIMSPGSLPSLHVRKQKALELDAQHLSETFDNIDNLS 780

QY 781 PKASHRSQKHQSLYGDYVDTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSSRGS 840
 Db 781 PKASHRSQKHQSLYGDYVDTNRHDDNRDNFTGNMTVLSPLYNTTVLPSSSSSRGS 840

QY 841 LSSRSEKDRSLERERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900
 Db 841 LSSRSEKDRSLERERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAIHTS 900

QY 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNYNTFKSENSNRTCSMPYAKLEYKRSS 960
 Db 901 QEDRSSGTTLHCVTDERNALRSSAAHTSNYNTFKSENSNRTCSMPYAKLEYKRSS 960

QY 961 NDSLNSVSSSDGYKRGQMKPSIESYEDDSKFCYGOYPADLAHKITHSANHMDNDGE 1020
 Db 961 NDSLNSVSSSDGYKRGQMKPSIESYEDDSKFCYGOYPADLAHKITHSANHMDNDGE 1020

QY 1021 LDTPIYSLKYSDQLNSGRQSPSONERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080
 Db 1021 LDTPIYSLKYSDQLNSGRQSPSONERWARPKHIIIEDEIKQSEQRQSRNQSTTYPVYTE 1080

QY 1081 STDDKHLKFQPHFGQECVSPYRSRGANGSETNRVSGNHGINQVNSOSLQCEDDYEDDKP 1140
 Db 1081 STDDKHLKFQPHFGQECVSPYRSRGANGSETNRVSGNHGINQVNSOSLQCEDDYEDDKP 1140

QY 1141 TNSYERYSEBEOHEEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKSFSPFSKS 1200
 Db 1141 TNSYERYSEBEOHEEBERPTNYSIKYNEEKRHVDQPIDYSLKYATDIPSSQKSFSPFSKS 1200

QY 1201 SSGOSSKTEHMSSESSENTSTPSSNAKRONLHPSSAQSRGQPKAATCKVSSINOETIQ 1260
 Db 1201 SSGOSSKTEHMSSESSENTSTPSSNAKRONLHPSSAQSRGQPKAATCKVSSINOETIQ 1260

QY 1261 TYCVEDTPICFSRCSSLSLSSAEDBIGCQNTTQEZADSANTLQIAIEKIKGIRSAEDPV 1320
 Db 1261 TYCVEDTPICFSRCSSLSLSSAEDBIGCQNTTQEZADSANTLQIAIEKIKGIRSAEDPV 1320

QY 1321 SEVPAYSOHPRTKSRLOQSSLSSESARHKAVFFSSGAKSPSKSGAQTPKSPPEHYVQET 1380
 Db 1321 SEVPAYSOHPRTKSRLOQSSLSSESARHKAVFFSSGAKSPSKSGAQTPKSPPEHYVQET 1380

QY 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGOTMPPSRSKTTP 1440
 Db 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGOTMPPSRSKTTP 1440

QY 1441 PPQTQATKREVPKNKAPTAEKRESGPKQAAVNAAVQVRQVLPDADTLHFAESTPDGF 1500
 Db 1441 PPQTQATKREVPKNKAPTAEKRESGPKQAAVNAAVQVRQVLPDADTLHFAESTPDGF 1500

QY 1501 SCSSLSALSALDLPFFTKQVELRIMPVQENDNGNETESEQKESNENOEKEAETIDSE 1560
 Db 1501 SCSSLSALSALDLPFFTKQVELRIMPVQENDNGNETESEQKESNENOEKEAETIDSE 1560

QY 1561 KOLLDDSDDDDIIEBEECIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
 Db 1561 KOLLDDSDDDDIIEBEECIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620

QY 1621 QNRLOQKQVSTFTPGDMPRVYCVETGPIINFSTATSLDLTITSPNEIAGAEGVGGQA 1680
 Db 1621 QNRLOQKQVSTFTPGDMPRVYCVETGPIINFSTATSLDLTITSPNEIAGAEGVGGQA 1680

QY 1681 SGFEKXRTIPTEGRSTDEAOGGKTSVTIPELDDNKAEBEGDILAEICINSAMPKGSHP 1740
 Db 1681 SGFEKXRTIPTEGRSTDEAOGGKTSVTIPELDDNKAEBEGDILAEICINSAMPKGSHP 1740

QY 1741 FRVKKIMDVQOQASASSAPNKNQLDGKKKPTSPVKPIPONTETRYTRVKNADSKNNLN 1800
 Db 1741 FRVKKIMDVQOQASASSAPNKNQLDGKKKPTSPVKPIPONTETRYTRVKNADSKNNLN 1800

QY 1801 AERVFSNKNDSKQNLKNNKDFNDKLPNNEDRVGSPAFDGPHTPTIEGTPYCFERN 1860
 Db 1801 AERVFSNKNDSKQNLKNNKDFNDKLPNNEDRVGSPAFDGPHTPTIEGTPYCFERN 1860

QY 1861 SLSSLDPDDDDVDLSREKAEELKAKENKESAKVTSHTSLTNSQOSANKTQIAKQPINR 1920
 Db 1861 SLSSLDPDDDDVDLSREKAEELKAKENKESAKVTSHTSLTNSQOSANKTQIAKQPINR 1920

QY 1921 GQPKPILOKQSTFPQSSKDIIDRGAAATDEKLNFAIENTPVCFSENSSLSLSDIDQENN 1980
 Db 1921 GQPKPILOKQSTFPQSSKDIIDRGAAATDEKLNFAIENTPVCFSENSSLSLSDIDQENN 1980

QY 1981 NKENEPKETEPPDSDGCEPSKPAQAGYAPKSFVSDTTPVCFGRNSLSLSLSDSDDLQ 2040
 Db 1981 NKENEPKETEPPDSDGCEPSKPAQAGYAPKSFVSDTTPVCFGRNSLSLSLSDSDDLQ 2040

QY 2041 ECISAMPKKKPSRLKGNKHSRNNKGIIGEDLTLDKDIQRPDSEHGLSPSENFD 2100
 Db 2041 ECISAMPKKKPSRLKGNKHSRNNKGIIGEDLTLDKDIQRPDSEHGLSPSENFD 2100

QY 2101 WKAIQEGANSIVSSLHQAAAAACLSRQASSDSLSLSKSGISLGSFPFHLTPDQEKPT 2160

Db 2101 WKAIQEGANSIVSSLSHQAACAAACLSRQSSSDSLSLKSIGSLGSPHLLTPDQEKPT 2160
 QY 2161 SNKGRPILPGEKSTLTETTKIIESEKIGKGGKVKVSLITGKVRNSSETSGQMKOPLOAN 2220
 Db 2161 SNKGRPILPGEKSTLTETTKIIESEKIGKGGKVKVSLITGKVRNSSETSGQMKOPLOAN 2220
 QY 2221 MPSISGRMTWHPGVNRSSSTSPVSKGKPLKTPAKSPSEGGTATTPRGAKPVSYS 2280
 Db 2221 MPSISGRMTWHPGVNRSSSTSPVSKGKPLKTPAKSPSEGGTATTPRGAKPVSYS 2280
 QY 2281 ELSFVARQTSQIGGSKAPRSRGRDSTPSRAQPLSRPIQSPGRNLSIPRNGISPPN 2340
 Db 2281 ELSFVARQTSQIGGSKAPRSRGRDSTPSRAQPLSRPIQSPGRNLSIPRNGISPPN 2340
 QY 2341 KLSQLPRTSPSTASTKSSGGKMSYTPSPQMSQQNLTKQGLSKNASSIPRSESASG 2400
 Db 2341 KLSQLPRTSPSTASTKSSGGKMSYTPSPQMSQQNLTKQGLSKNASSIPRSESASG 2400
 QY 2401 LQNMNNGANKVELSRMSSTKSSGSDRSPVLRQSTFIKEAPSTLRKLESA 2460
 Db 2401 LQNMNNGANKVELSRMSSTKSSGSDRSPVLRQSTFIKEAPSTLRKLESA 2460
 QY 2461 SFESLSPSPSPASPTQSOATPVLSPLDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 Db 2461 SFESLSPSPSPASPTQSOATPVLSPLDMSLSTHSSVQAGGWRKLPNLSPTIEYNDG 2520
 QY 2521 RPAKRDHIAHSHSPSLRPNISGRTWKRHSKSSSLPRVSTWRTTGSSSSILGASBS 2580
 Db 2521 RPAKRDHIAHSHSPSLRPNISGRTWKRHSKSSSLPRVSTWRTTGSSSSILGASBS 2580
 QY 2581 SEKAKSEDEKHVNSISGTSKOSKNOVSAGTKWIKENEFSPNNSQTVSSGATNGAES 2640
 Db 2581 SEKAKSEDEKHVNSISGTSKOSKNOVSAGTKWIKENEFSPNNSQTVSSGATNGAES 2640
 QY 2641 KTLIIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 Db 2641 KTLIIYQAPAVSKTEDVWVRIEDCPINNPRSGRPTGNTPPVIDSVSEKANPNIKDSKN 2700
 QY 2701 QAKQNGVGVMPRTVGLNRLNSFTQVADPQKTEIKPQNNPVPVSETNESSIVERT 2760
 Db 2701 QAKQNGVGVMPRTVGLNRLNSFTQVADPQKTEIKPQNNPVPVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSPSGTVAARVTFNNPSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
 Db 2761 PFSSSSSKHSPSGTVAARVTFNNPSPRKSADSTSGARPSQIPTPVNNNTKRDST 2820
 QY 2821 DSTESSGTQSPKRHSGLYVTSV 2843
 Db 2821 DSTESSGTQSPKRHSGLYVTSV 2843

RESULT 2
 APC_RAT
 ID APC_RAT STANDARD; PRT; 2842 AA.
 AC P70478;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein).
 GN APC.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP STRAIN=Fischer 344/N; TISSUE=Brain;
 RX MEDLINE=96116966; PubMed=8563176;
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,
 RA Sugimura T., Nagao M.;
 RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
 RL Mamm. Genome 6:746-748(1995).

[2]
 MUTAGENESIS.
 RC STRAIN=Sprague-Dawley, and Fischer 344/N;
 RX MEDLINE=95148647; PubMed=7846077;
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
 RA Weisburger J.H., Sugimura T., Nagao M.;
 RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
 tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
 CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 and participates in Wnt signaling. APC activity is correlated with
 its phosphorylation state (By similarity).
 CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
 axin (By similarity).
 CC -!- PTM: Phosphorylated by GSK3B (By similarity).
 CC -!- SIMILARITY: Contains 7 ARM repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D38629; BAA07609.1; -;
 DR HSP; Q02248; 3BCT.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 4.
 DR SMART; SMO0185; ARM; 5.
 DR PROSITE; PS0176; ARM_REPEAT; 1.
 KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
 Repeat.
 FT DOMAIN 1 728 LEU-RICH.
 FT DOMAIN 125 260 COILED COIL (POTENTIAL).
 FT REPEAT 451 493 COILED COIL (POTENTIAL).
 FT REPEAT 503 545 ARM 1.
 FT REPEAT 546 589 ARM 2.
 FT REPEAT 590 636 ARM 3.
 FT REPEAT 637 681 ARM 4.
 FT REPEAT 682 723 ARM 5.
 FT REPEAT 724 765 ARM 6.
 FT REPEAT 739 783 ARM 7.
 FT DOMAIN 1130 1155 SER-RICH.
 FT DOMAIN 1356 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 ASP/GLU-RICH (ACIDIC).
 FT MUTAGEN C->R: IN AN IQ-INDUCED COLON TUMOR.
 SQ SEQUENCE 2842 AA; 310530 MW; 3CB2EA8A34E8F47 CRC64;
 Query Match 90.6%; Score 13198.5; DB 1; Length 2842;
 Best Local Similarity 90.0%; Pred. No. 0;
 Matches 2563; Conservative 121; Mismatches 154; Indels 9; Gaps 8;
 QY 1 MAASVDQLKQVEALKMENSRLQELDNSNHLTKLETSANNKEVLKOLQGSIEDAM 60
 Db 1 MAASVDQLKQVEALKMENSRLQELDNSNHLTKLETSANNKEVLKOLQGSIEDAM 60
 QY 61 ASSGQIDLLERLKEINLDSNFPFGVKLRSMVLSYSGRSGSVSRSGECSFVPMGSPFR 120
 Db 61 -TSQIDLLERLKEINLDSNFPFGVKLRSMVLSYSGRSGSVSRSGECSFVPMGSPFR 118
 QY 121 RGFVNGSRSTGYLEELKEKSLALLADKEKEDWYVYAOQLNLTKRIDSPLTENFSL 180
 Db 119 RAVFNGSRSTGYLEELKEKSLALLADKEKEDWYVYAOQLNLTKRIDSPLTENFSL 178
 QY 181 QDTMTFRQLEYEARQIRVAMEEQLGTCDMEKRAQRRIARIQQIEKDIRLROLQSOAT 240
 Db 179 QDTMTFRQLEYEARQIRVAMEEQLGTCDMEKRAQRRIARIQQIEKDIRLROLQSOAA 238
 QY 241 EAERSSQNHGTGSHDAERQNEGVGVGINMATSGNGSGSTTRMDHETASVLSSTHSA 300

Db 239 EAERSQKHETASHEAEQLEGGQVAESNLATSGGSSAAARVDHETAGVLSSSGTHSA 298
Qy 301 PRLTSHLGTKVEMYSLLSMLGTHDKDMSRTLLAMSSODSCISMOSGCLPILLIQL 360
Db 299 PRLTSHLGTKVEMYSLLSMLGTHDKDMSRTLLAMSSODSCISMOSGCLPILLIQL 358
Qy 361 HGNDDKSVLLGNSRSGKEARASAAALHNIHISQDDPKRGRREIRVLHLLBOIRAYCETC 420
Db 359 HGNDDKSVLLGNSRSGKEARASAAALHNIHISQDDPKRGRREIRVLHLLBOIRAYCETC 418
Qy 421 WEQZAEHPEGMDDKNPMPAPVEHQCIPAVCVLMKLSFDEEHRHAMNELGGLOIAEALLO 480
Db 419 WEQZAEHPEGMDDKNPMPAPVEHQCIPAVCVLMKLSFDEEHRHAMNELGGLOIAEALLO 478
Qy 481 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLSESDL 540
Db 479 VDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLSESDL 538
Qy 541 QOVIASVLNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSAHCT 600
Db 539 QOVIASVLNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVLNLSAHCT 598
Qy 601 ENKADICAVDGAFLVGTLYRSQTNLTALIESGGILRNVSLLIATNEDHRQLRENN 660
Db 599 ENKADICAVDGAFLVGTLYRSQTNLTALIESGGILRNVSLLIATNEDHRQLRENN 658
Qy 661 CLQTLQHLKSHSLTIVSNACOTLNLNLSARPDKQDALWDGAVSMKXNLHSHKXMIAM 720
Db 659 CLQTLQHLKSHSLTIVSNACOTLNLNLSARPDKQDALWDGAVSMKXNLHSHKXMIAM 718
Qy 721 GSAALRNLMANRPKAKYKDNIMSPGSLPSLHVKKQKALAEALDAQHLSETFDNIDLS 780
Db 719 GSAALRNLMANRPKAKYKDNIMSPGSLPSLHVKKQKALAEALDAQHLSETFDNIDLS 778
Qy 781 PKASHRSQRKHQSUYGVYFTNDRDNRDNFNNTGNMVLSPYLNLTVPSSSSSRGS 840
Db 779 PKASHRSQRKHQSUYGVYFTNDRDNRDNFNNTGNMVLSPYLNLTVPSSSSSRGS 838
Qy 841 LDSSRSKDRSLERBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAHLS 900
Db 839 LDSSRSKDRSLERBERGIGLGNYPATENPGTSSKRGLOISTTAAQIAKWEEVSAHLS 898
Qy 901 QEDRSSGTTLEHCVTDERNALRRSSAAHTSNYNTFTKSENSNRCTCMYPKALEYKRSS 960
Db 899 QEDRSSGTTLEHCVTDERNALRRSSAAHTSNYNTFTKSENSNRCTCMYPKALEYKRSS 958
Qy 961 NDLSNVSSDGYGKRGOMKPSIEYSDDSKFCYQYPADLAHKHISANHMDNDGE 1020
Db 959 NDLSNVSSDGYGKRGOMKPSIEYSDDSKFCYQYPADLAHKHISANHMDNDGE 1018
Qy 1021 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1080
Db 1019 LDTPIYSLKYSDEQLNSGRQSPSONERWARPKHIEDEIKQSEORQSRNOSTTYPVYTE 1078
Qy 1081 STDDKHLKFPQHFQOQECVPSYRSGANGSETRNVGSHNGINQNVOSLCOEDDYEDDKP 1140
Db 1079 NTDDKHLKFPQHFQOQECVPSYRSGANGSETRNVGSHNGINQNVOSLCOEDDYEDDKP 1138
Qy 1141 TNYSERVSEEOHEEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1200
Db 1139 TNYSERVSEEOHEEERPTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1198
Qy 1201 SSGQSKTEHSSSSSENTSTPSSNAKRONQLHPSSAQSRGQPOKAATCKVSSINQETIQ 1260
Db 1199 PSVQTKTEHNSPSSSEAAASAPSSNAKROSLHPSSAQ-RNQTPKGTACKVPSINQETMQ 1257
Qy 1261 TYCVEDTIPICSRCSLSLSAEDIEGNCOTTOEADSANTLOIAEIKEKIGTISAEDPV 1320
Db 1258 TYCVEDTIPICSRCSLSLSAEDIEGNCOTTOEADSANTLOIAEIKEKIGTISAEDPV 1317
Qy 1321 SEVPAVSQHPRTKSRRLQSGSLSSSARHKAVERFSSGAKSPSKGAOTPKSPPEHYQET 1380
Db 1318 SDVPAVSQSTRPKSRLOASGLASEARHKAVERFSSGAKSPSKGAOTPKSPPEHYQET 1377

Qy 1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGOTMPSRSKT-- 1438
Db 1378 PLVFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGOTMPSRSKTTP 1437
Qy 1439 PPPPQTAQTKREVPKVAKAPTAEKRSBGPKQAAVANAQVQVLPDADTLLHFATESTPD 1498
Db 1438 PPPPQVQTKREVPKVAPAEQEGGPKQTAASAQVQVLPDADTLLHFATESTPD 1497
Qy 1499 GFSCSSLSALSJDEPFIOKDVLRMPVQENDNGNETESEOPKESNENQKEAEKTIID 1558
Db 1498 GFSCSSLSALSJDEPFIOKDVLRMPVQENDNGNETESEOPKESNENQKEAEKPD-D 1556
Qy 1559 SEKOLLDDSDDDIEILEECIIISAMPTKSSRKAQKPAQTASKLPPPVARFSPQLPVYKLL 1618
Db 1557 SEKOLLDDSDDDIEILEECIIISAMPTKSSRKAQKPAQTASKLPPPVARFSPQLPVYKLL 1616
Qy 1619 PSQNELPOKHVSFTPGDDMPRVYCVBGTPIINSTATSLSDLIIESPPNELAAGEVRGG 1678
Db 1617 PSQSRLOAQKHVSFTPGDDMPRVYCVBGTPIINSTATSLSDLIIESPPNELAAGEVGRAS 1676
Qy 1679 AQGEPEKRTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEEDGIIAECINSAMPKGS 1738
Db 1677 VQSEPEKRTIIPTEGRSTDEAOGGKTSSVTIPELDDNKAEEDGIIAECINSALPKORSH 1736
Qy 1739 KPRVKKIMQVQOASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKNN 1798
Db 1737 KPRVKKIMQVQOASASSAPNKNQLDGKKKPTSPVKPIPONTYRTRVRKNADSKNN 1796
Qy 1799 LNAERFSDNKNKSKQNLKNNKDFDKLPNNEDRVGRSFAFDPHHYPIEGTVPYCFSR 1858
Db 1797 VNTTEFSDNKNKSKQNLKNNKDFDKLPNNEDRVGRSFAFDPHHYPIEGTVPYCFSR 1856
Qy 1859 NDLSLSDPDDDDVLSREKAEELKAKENKESAKVTSHTELTSSNOQSANKTOAIKQPI 1918
Db 1857 NDLSLSDPDDDDVLSREKAEELKAKENKESAKVTSHTELTSSNOQSANKTOAIKQPI 1916
Qy 1919 NRQOPKILQKQSTFPQSSKDIIDRGAATDEKLONFAIENTPVCFSHNSLSLSLSDIDQE 1978
Db 1917 NRQOPKILQKQSTFPQSSKDIIDRGAATDEKLONFAIENTPVCFSHNSLSLSLSDIDQE 1976
Qy 1979 -NNKNENPIKETEPDPSQGEPSKPAQSGVAPKSFHVEDTPVCFSRNSSLSSLSIDSEDD 2037
Db 1977 NNNNEETGPVDAEPANAOQPGKPAQSGVAPKSFHVEDTPVCFSRNSSLSSLSIDSEDD 2036
Qy 2038 LLOECISSAMPKPKPSRLKGDNEKHSRNMGGILGEDTLDKIQRDSEHGLSPDSE 2097
Db 2037 LLOECISSAMPKPKPSRLKGDNEKHSRNMGGILGEDTLDKIQRDSEHGLSPDSE 2096
Qy 2098 NFDWKAIQEGANSIVSLSLHO-AAAAACLROASSDSLSLSKSGISLGSPPHLLTPDQEE 2156
Db 2097 NFDWKAIQEGANSIVSLSLHO-AAAAACLROASSDSLSLSKSGISLGSPPHLLTPDQEE 2156
Qy 2157 KPFTSNKGRPILKPGPKSTLETKKIESGKIGKGVKYSLTGKVRNSENSEISGOMKQ 2216
Db 2157 KPFTSNKGRPILKPGPKSTLETKKIESGKIGKGVKYSLTGKVRNSENSEISGOMKQ 2216
Qy 2217 LOANMPSISRGRTMIH PGVRNSSLSSSTSPVSKGPPKTPASKSPSEGOATTPSPGAKP 2276
Db 2217 LOANMPSISRGRTMIH PGVRNSSLSSSTSPVSKGPPKTPASKSPSEGOATTPSPGAKP 2276
Qy 2277 SVKSELSPVARQTSQIGGSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNGLSPGRNGI 2336
Db 2277 AVKSELSPIRQTSIHISGNGKPSRSGSRDSTPSRPTQPLSRPMQSPGRNGLSPGRNGI 2336
Qy 2337 SPNKLSQLPRTSSPSTASTKSGSGKMSYTPSGROMSQOONLTQOTGLSKNASSIPRSES 2396
Db 2337 STPNKLSQLPRTSSPSTASTKSGSGKMSYTPSGROMSQOONLTQOTGLSKNASSIPRSES 2396
Qy 2397 ASKGLNOMNNGANGKVKVELSRMSSTKSSGSSDRSERPVLVRSQSTFIKEAPSPTLRRKL 2456
Db 2397 ASKGLNOMNNGANGKVKVELSRMSSTKSSGSSDRSERPVLVRSQSTFIKEAPSPTLRRKL 2456

QY	2457	BESASPELSPSRPASPTRSQOQTVLSPSLPDMSLSTHSSVQAGWKLPPNLSPTIE	2516
DB	2457	BESASPELSPSRPASPTRSQOQTVLSPSLPDMSLSTHSSVQAGWKLPPNLSPTIE	2516
QY	2517	YNDGRPAKRDHDIARSHSESPSRIPINRSQGTWREKSHSSLPVSTWRTGSSSSILSA	2576
DB	2517	YSDGRSKRDHDIARSHSESPSRIPVNRAGTWREKSHSSLPVSTWRTGSSSSILSA	2576
QY	2577	SSSESEKAKSEDKHNSISGTVKQSNQVSAKGTWRKIKENEFSTNSTSTQTVSSGATN	2636
DB	2577	SSSESEKAKSEDKHNSVPGPQMKENQVPTGTWRKIKESBITNTVSTQTVSSGAAS	2636
QY	2637	GASKTLIYQMAVAKSTEDVWVRIEDCPINPRSGRSPGTNTPPVIDSVSKANPNIKD	2696
DB	2637	GASKTLIYQMAVAKSTEDVWVRIEDCPINPRSGRSPGTNTPPVIDSISEKGNPSIKD	2696
QY	2697	SKNQAKQNVGSGVPMRTVGLNRLNLSFIQVADPOKGTETKPGQNPVPVSEINNESSI	2756
DB	2697	SKDTQKQSGVSGSGS-PVQTVGLNRLNLSFIQVAPSKGTETKAGQSGPAPVAETGETCM	2755
QY	2757	VERTPSSSSSSKHSSPSGTVAAVTPFNPNPRKSSADTSARPSQIPTPVNNNTKKR	2816
DB	2756	AERTPSSSSSSKHSSPSGTVAAVTPFNPNPRKSSADTSARPSQIPTPVGSSSTKKR	2815
QY	2817	DSKTDSTESSGTSQPKRHSGSYLTVSV	2843
DB	2816	DSKTDSTESSGAQSPKRHSGSYLTVSV	2842
RESULT 3			
APC_MOUSE	ID	APC_MOUSE	
AC	Q61315	Q62044	STANDARD; PRT; 2845 AA.
DT	16-OCT-2001	(Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	Adenomatous polyposis coli protein (APC protein) (mAPC).		
GN	APC		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.		
RC	STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;		
RX	MEDLINE=92263101; PubMed=1350108;		
RA	Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,		
RA	Luongo C., Gould K.A., Dove W.F.;		
ET	"Multiple intestinal neoplasia caused by a mutation in the murine		
RT	homolog of the APC gene."		
RL	Science 256:668-670(1992).		
RN	[2]		
RP	ERRATUM.		
RA	Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,		
RA	Luongo C., Gould K.A., Dove W.F.;		
RL	Science 256:1114-1114(1992).		
RN	[3]		
RP	SEQUENCE OF 1-45 FROM N.A.		
RC	STRAIN=BALB/c; TISSUE=Liver;		
RA	Dicker F., Lambertz S., Reitmair A., Ballhausen W.G.;		
RT	"The murine APC gene: alternative splicing of 5' untranslated		
RT	region segments."		
RL	Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	ALTERNATIVE SPLICING.		
RX	MEDLINE=94061824; PubMed=8242607;		
RA	Oshima M., Sugiyama H., Kitagawa K., Taketo M.;		
RT	"APC gene messenger RNA: novel isoforms that lack exon 7.;"		
RL	Cancer Res. 53:5589-5591(1993).		
CC	-!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1		
CC	and participates in Wnt signaling. APC activity is correlated with		
CC	its phosphorylation state (By similarity).		
CC	-!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds		

CC	axin (By similarity).		
CC	-!- ALTERNATIVE PRODUCTS:		
CC	Event=Alternative splicing; Named isoforms=4;		
CC	Name=1;		
CC	Isoid=Q61315-1; Sequence=Displayed;		
CC	Name=2;		
CC	Isoid=Q61315-2; Sequence=VSP_004116;		
CC	Name=3;		
CC	Isoid=Q61315-3; Sequence=VSP_004117;		
CC	Name=4;		
CC	Isoid=Q61315-4; Sequence=VSP_004116, VSP_004117;		
CC	-!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,		
CC	lung, brain, stomach, intestine, testis and ovary.		
CC	-!- PTM: Phosphorylated by GSK3B (By similarity).		
CC	-!- SIMILARITY: Contains 7 ARM repeats.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL; M88127; AB59632.1; -;		
DR	EMBL; U02937; AA03443.1; -;		
DR	PIR; I49505; I49505.		
DR	HSP; Q02248; 3BCT.		
DR	MGI; MGI:88039; Apc.		
DR	GO; GO:0005737; C:cytoplasm; IDA.		
DR	GO; GO:0005634; C:nucleus; IDA.		
DR	GO; GO:0080133; F:beta-catenin binding; IDA.		
DR	GO; GO:0009952; P:anterior/posterior pattern formation; IMP.		
DR	GO; GO:0009798; P:axis specification; IMP.		
DR	GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.		
DR	GO; GO:0016053; P:Wnt receptor signaling pathway; IDA.		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR00225; Armadillo.		
DR	Pfam; PF00514; Armadillo_seg; 4.		
DR	SMART; SM00185; ARM; 5.		
DR	PROSITE; PS0176; ARM_REPEAT; 1.		
DR	Wnt signaling pathway; Anti-oncogene; Phosphorylation;		
KW	Alternative splicing; Repeat; Coiled coil.		
FT	DOMAIN 1 61		
FT	DOMAIN 125 245		
FT	DOMAIN 1 728		
FT	REPEAT 451 493		
FT	REPEAT 503 545		
FT	REPEAT 546 589		
FT	REPEAT 590 636		
FT	REPEAT 637 681		
FT	REPEAT 682 723		
FT	REPEAT 724 765		
FT	DOMAIN 739 2834		
FT	DOMAIN 1130 1156		
FT	DOMAIN 1556 1575		
FT	DOMAIN 1864 1891		
FT	VARSPPLIC 243 276		
FT	VARSPPLIC 310 410		
FT	VARSPPLIC 120 120		
FT	VARIANT 493 493		
FT	VARIANT 797 797		
FT	VARIANT 1330 1330		
FT	VARIANT 1618 1618		
FT	VARIANT 2294 2294		
FT	VARIANT 2496 2496		
FT	VARIANT 2523 2523		
FT	VARIANT 2813 2813		
FT	SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;		
SQ			
Query Match 90.5%; Score 13178; DB 1; Length 2845;			

RA Chen F., Rebay I.;
 RT "split ends, a new component of the Drosophila EGF receptor pathway,
 RL regulates development of midline glial cells.";
 RN Curr. Biol. 10:943-946(2000).
 [9]
 RP FUNCTION ON WG PATHWAY.
 RX MEDLINE=22668876; PubMed=12783785;
 RA Lin H.V., Doroquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;
 RT "Split ends is a tissue/promoter specific regulator of Wingless
 RL signaling.";
 CC Development 130:3125-3135(2003).
 CC -!- FUNCTION: Probable corepressor protein, which regulates different
 CC key pathways such as the EGF receptor and Wg pathways. Involved in
 CC neuronal cell fate, survival and axon guidance, cell cycle
 CC regulation and repression of head identity in the embryonic trunk.
 CC May act with the Hox gene Deformed and the EGF receptor signaling
 CC pathway. Positive regulator of the Wg pathway in larval tissues
 CC but not in embryonic tissues. May act as a transcriptional
 CC corepressor protein, which repress transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative promoter;
 CC Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
 CC alternative promoters;
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1;
 CC IsoId=Q8SX83-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8SX83-2; Sequence=VSP_008565, VSP_008566;
 CC Note=No experimental confirmation available;
 CC Name=3; Synonyms=SpnL;
 CC IsoId=Q8SX83-3; Sequence=VSP_008567;
 CC Note=Produced by alternative splicing of isoform 1;
 CC Name=4; Synonyms=SpnS;
 CC IsoId=Q8SX83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
 CC Note=Produced by alternative splicing of isoform 2;
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
 CC in stage 3 embryos, and in blastoderm cells, including pole cells.
 CC Expressed throughout the rest of embryogenesis. Later, it is
 CC expressed at higher level in epidermal cells and CNS.
 CC -!- DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
 CC zygotically.
 CC -!- SIMILARITY: Belongs to the Spn family.
 CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
 CC -!- SIMILARITY: Contains 1 SPOC domain.
 CC -!- CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF188205; AAF13218.1; -;
 CC EMBL; AF184612; AAF26299.1; -;
 CC EMBL; AF221715; AAF34661.1; ALT_INIT.
 CC EMBL; AE003590; AAF51534.2; -;
 CC EMBL; AE003590; AAF51535.2; -;
 CC EMBL; AE003590; AAN10511.1; -;
 CC EMBL; AY094788; AAM11141.1; ALT_SEQ.
 CC HSP; P09551; IHA1.
 CC FlyBase; FBgn0016977; spen.
 CC DR GO; GO:0007411; P:axon guidance; IMP.
 CC DR GO; GO:0008347; P:glia cell migration; IMP.
 CC DR InterPro; IPR000504; RNA_rec_mot.
 CC DR Pfam; PF00076; rrm; 3.
 CC DR SMART; SM00360; RRM; 3.
 CC DR PROSITE; PS0102; RRM; 3.
 CC DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.

DR PROSITE; PS00917; SPOC; 1.
 KW Transcription regulation; Repressor; Developmental protein;
 KW Nuclear protein; Repeat; RNA-binding; Coiled coil;
 KW Alternative promoter usage; Alternative splicing.
 FT DOMAIN 554 632 RNA-BINDING (RRM) 1.
 FT DOMAIN 656 730 RNA-BINDING (RRM) 2.
 FT DOMAIN 734 806 RNA-BINDING (RRM) 3.
 Query Match 3.3%; Score 474; DB 1; Length 5560;
 Best Local Similarity 18.5%; Pred. No. 8.3e-10;
 Matches 621; Conservative 479; Mismatches 1219; Indels 1034; Gaps 147;
 QY 2 AASYDQLLKQVE--ALKWENSLRQELSDNSHLTKLETASNKKVILKQSGIEDE 58
 Db :::::STASVTASISISSSTNGNSLTSAIHVQKFCQSTFVEEHTKSGTSTQS 1845
 QY 59 AMASSQIDLLERKELNLDSSNFGVKLRKMSLRSGYSGVSSRSGECSVPVMSGF 118
 Db :::::SSSSKKI-----SSTHD--KLHSHKNNRSESDKKIKKSDKNASSDKKNSS 1891
 QY 119 PRGFVNGRSTGYLEE-----LEKERSLLADLDKEEK 153
 Db 1892 TSQ-----SSKSATPRIEDDSSEADDTADKAENQKHEKKEKREKDLKQVREK 1947
 QY 154 EKDWYIAQNLTKRIDSPLTENFSLQDMTRQLYEARQIRVAMEBQLGTCQDWEKR 213
 Db 1948 DRK--AQEEREK-----EDRAKEBEKEREREKKAQEDREKEREERELREKQ 1996
 QY 214 -----AQRIARIQIQEKD-----ILRIQLQSQATEAERSQNH-ETGSHDAER 259
 Db 1997 DKEQKEIREKDLREKQERDRNREKELDKLREKREKEKREKELHREKQDREH 2056
 QY 260 ONEGGQVGVEINMATSGNGOGSTTRMDHETASVLSSTSTHSAPRRLTSLHGTQVEM---VY 316
 Db 2057 REKEQRRAMDVEQERG-----GRM-----RELSSYQSKMDIAGEAS 2095
 QY 317 SLISMLGTHDKD-----DMSRTLLAMS-----SSQDS 343
 Db 2096 SLTAIDCQHNKENAMDTIAQTGASPTSDNTPKERSKLSRNSPVLLHKKRLSSQS 2155
 QY 344 CISMRQSGCLPLLIQLHNGD-----KDSVLLGNRSGKEAR-----380
 Db 2156 NBSAGGGSGGSGSHQIHEDYVYKIRIMNSQNI SVHSSNQRLNDRRDSKEHKSSTFKED 2215
 QY 381 -----ARASAAALNITHSQPDQKGRREIR-----VLHLEQIRAYCE 418
 Db 2216 KNSSSHIRPHGCGSSASSSKHHRRDKHQKGSASSIETNWSIEVVVDPISTQKHNLN 2275
 QY 419 TCWEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHHRHAMMELGLOIAEL 478
 Db 2276 TSEELQSHQPKREKER-----EHFSGHANSSSSRHKSXKIDHHHREK--KRHSVAES 2326
 QY 479 LQVDCMYGLT-NDHYSITLRRYAGWALTNLFGDVANKATLCMKGCMRALVAQLKSES 537
 Db 2327 TWTDEHTTQQHNPHRIS-----AAGSGAGELSSAATNTS-----2363
 QY 538 EDLQOVIAVLRLNLSWRADVNGKTLREVGSKVLMCALEKVEKSTLKVLSALNLSA 597
 Db 2364 -----SGKLHHQHRRSVVERKSSR---GS-----DEGHHSSSKSLRAKMLMSS 2404
 QY 598 HCTENKADICAVDGAFLVGLITYESQNTLAIIBSGGGILRNVSLLATNEDHRIQLR 657
 Db 2405 -----ADSDDTDDA-----2413
 QY 658 ENNCLOTLLQHLKSHSLTIVSNACGTLNLSARNPKDQBALMDGAVSMLKNL-----710
 Db 2414 -----SKKHSIFDIPDDCPNV-----SMYDKVKARCKMQRQAEEK 2450
 QY 711 -HSHKHOMIAMSAAALNLM--ANRPAYKQANIMSPGSSLPPLHLVR-----755
 Db 2451 KIKAKTSQLKQSPAKKKTSTSYDGSDDTEFEFQHRNNGSS--SPHGYPGLSSSSDDDD 2508
 QY 756 ---KQALEAEALDAQHLSETFDNIDNLSPKASHRSKQRH-----K 792

QY 2630 VSSGATNGAESKTLIYQMAVASKTEDVWRIEDC--PINNPSRGRPT-GNTPPVDSV 2686
 Db ISSVAPSAEDQRP--RLILTINKTQPSIKNISEMEQTIOQQOQQQSEVINTDPIGGDN 4496
 QY 2687 SEKANPNIKSKDQAKQK--VNGSVPMRTVGLNRLNSFIQVDAPD-QK 2734
 Db 4497 SESC--NTRKSRRLQEKEDRSTVDIIEDVVVNTNTPTGT-----GPHLPK 4540
 QY 2735 GTEIKP--GQN-----NPV-----PVSETNESSIVERTPPSSSSSSK 2769
 Db 4541 GAOPTPRSGRGAQAKTDAVQIINAVGRPRSRDKRTIGTQNTANLIEEVTASNAATVAAS 4600
 QY 2770 HSSFSGTVAR-----VTPNPNPSRKSSADSTSRPQIP 2806
 Db 4601 HLAPEGAGVESHVPQLDAKEVPSVVVTPIS--TPAPVSAAPTVVPVPAWV 4652

RESULT 6
 MINT_HUMAN
 ID MINT_HUMAN STANDARD; PRT; 3664 AA.
 AC Q9GT58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
 GN MINT OR SHARP OR KIA0929.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
 RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
 RP RAR AND MTAIL1.
 RC TISSUE=Liver, and Pituitary;
 RX MEDLINE=21231190; PubMed=11331609;
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
 RA Hon M., Evans R.M.;
 RT "Sharp, an inducible cofactor that integrates nuclear receptor
 RT repression and activation.";
 RL Genes Dev. 15:1140-1151(2001).
 [2]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 294-3664 FROM N.A.
 RA Rhodes S., Huckle E.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
 RC TISSUE=Embryo, and Teratocarcinoma;
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otauki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE OF 2002-3664 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:163-70(1999).
 [6]
 RP INTERACTION WITH PPARD.
 RX MEDLINE=21874127; PubMed=11867749;

RA Shi Y., Hon M., Evans R.M.;
 RT "The peroxisome proliferator-activated receptor delta, an integrator
 RT of transcriptional repression and nuclear receptor signaling.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
 [7]
 RP FUNCTION, AND INTERACTION WITH RBPSUH.
 RX MEDLINE=2261914; PubMed=12374742;
 RA Oswald F., Koscewka U., Astrahantseff K., Bourtelee S., Dillinger K.,
 RA Zechner U., Ludwick L., Wilda M., Hameister H., Knoechel W., Lipray S.,
 RA Schmid R.M.;
 RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
 RT pathway.";
 RL EMBO J. 21:5417-5426(2002).
 [8]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
 RX MEDLINE=22777836; PubMed=12897056;
 RA Ariyoshi M., Schwabe J.W.R.;
 RT "A conserved structural motif reveals the essential transcriptional
 RT repression function of Spen proteins and their role in developmental
 RT signaling.";
 RL Genes Dev. 17:1909-1920(2003).
 CC -!- FUNCTION: Essential corepressor protein, which probably regulates
 CC different key pathways such as the Notch pathway. Negative
 CC regulator of the Notch pathway via its interaction with RBPSUH,
 CC which prevents the association between NOTCH1 and RBPSUH, and
 CC therefore suppresses the transactivation activity of Notch
 CC signaling. Blocks the differentiation of precursor B cells into
 CC marginal zone B cells. Probably represses transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins. May bind both to DNA and RNA.
 CC -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
 CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTAIL1. Interacts with
 CC RBPSUH; this interaction may prevent the interaction between
 CC RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and
 CC PPARD. Interacts with RAR in absence of ligand. Bind to the
 CC steroid receptor RNA coactivator SRA.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
 CC -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
 CC spleen and thymus. Expressed at intermediate level in kidney,
 CC liver, mammary gland and skin.
 CC -!- INDUCTION: By hormone 17-beta-estradiol (E2).
 CC -!- DOMAIN: The RID domain mediates the interaction with nuclear
 CC receptors (By similarity).
 CC -!- DOMAIN: The SPOC domain, which mediates the interaction with
 CC NCOR2, is essential for the repressive activity.
 CC -!- SIMILARITY: Belongs to the Spen family.
 CC -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.
 CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
 CC -!- SIMILARITY: Contains 1 SPOC domain.
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC EMBL; AF356524; AAK52750.1; -;
 DR EMBL; AL034555; CAB85442.1; ALT_SEQ.
 DR EMBL; AL034555; CAB85444.1; ALT_SEQ.
 DR EMBL; AL450998; -; NOT_ANNOTATED_CDS.
 DR EMBL; AL096858; CAB51072.1; ALT_INIT.
 DR EMBL; AK000882; BAA91405.1; ALT_INIT.
 DR EMBL; AK022949; BAB14324.1; ALT_INIT.
 DR EMBL; AB023146; BAA76773.1; -;
 DR InterPro; IPR000504; RNA_rec_mot.
 DR PDB; 1OW1; 19-AUG-03.
 DR Pfam; PF00076; rrm; 4.
 DR SMART; SM00360; RRM; 4.
 DR PROSITE; PS50102; RRM; 4.

DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 DR PROSITE; PS00917; SPOC; 1.
 KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;
 KW Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
 FT DOMAIN 1 573
 FT DOMAIN 6 81
 FT DOMAIN 335 415
 FT DOMAIN 438 513
 FT DOMAIN 517 589
 FT DOMAIN 688 715
 FT DOMAIN 977 1004
 FT DOMAIN 1170 1191
 FT DOMAIN 1408 1428
 FT DOMAIN 1496 1529
 FT DOMAIN 1592 1612
 FT DOMAIN 1928 1944
 FT DOMAIN 2201 2707
 FT DOMAIN 3498 3664
 FT DOMAIN 2130 2464
 FT DOMAIN 2709 2870
 FT DOMAIN 125 277
 FT DOMAIN 240 325
 FT DOMAIN 616 810
 FT DOMAIN 624 697
 FT DOMAIN 2428 2520
 FT DOMAIN 3220 3482
 FT VARIANT 970 970
 FT VARIANT 1091 1091
 FT VARIANT 2360 2360
 FT CONFLICT 956 956
 FT SEQUENCE 3664 AA; 402245 MW; 5228C58533E5B27B CRC64;

Query Match 3.1%; Score 452; DB 1; Length 3664;
 Best Local Similarity 19.3%; Pred. No. 3.3e-09;
 Matches 490; Conservative 341; Mismatches 969; Indels 742; Gaps 115;

QY 727 RNLAMRPAK-YKDNIMSPGSLPSLHVKKKALEAELDAQHSETFDNIDNLSPK-A 783
 DB 226 REYGRPRPENYCHSRSPHSS-----QSRNQSPQRLA 259
 QY 784 SHRSQRHQSLGYDVFDTNRHDDNRDNFNFTGNMVLSPYLNTVLPSSSSSSSGSLDS 843
 DB 260 SQASRPTRSPSGG-----SRSSSSSDS188-----SSSTSSSDSDS 297
 QY 844 SRSEKD-----RSLERE---RGIGLGNVHPATENPGTSSKRGLOI 880
 DB 298 SSSSSDSPARSVQSAVAPAPTSQLLSSLEKDEPKSGFKGVQNLVPRSTDSLKQGLFH 357
 QY 881 -----STTAAQIAKYMEVSAI--HTSQEDRSSSGSTT-----LHCVTD 917
 DB 358 EPKKFGKVTISQVHGTSERYGLVFRQDQEKALTASKGLFGQMIEVTAWIGPTE 417
 QY 918 ERNALR-----RSSAHT-----HSNTYNFTKS-----ENNRCTCMP- 950
 DB 418 SENEFPALDERIDEHPKATRLTFIGNLEKTTYTHDLNRNIFORGEIVDIDIKYNGVPQ 477
 QY 951 YAKLEY-----KRSDNLSNVSSSDGYGRGQMKPS-----IESYEDDESK--- 993
 DB 478 YAFLOQCDIASVCKAIKQWDGEYLGNNRLKLGFGK---SMPTNCVWLGLSSNVSDQYLT 534
 QY 994 --FCSTG-----QYPADLAHKIHSANHMDNDNGELDTPNYS-L 1029
 DB 535 RHFCRYGPVVKVVDRLKGMALVLYNEIY-AQAAVKETKGIKGNKIKVDFAKRESOL 593
 QY 1030 KYSDQELNSGRSPQNERWARPKHIEDEIKQSEORQS--RNQSTTY-----PVYTE 1080
 DB 594 AFVHCVEKSGQDIRDFYEMLAEE-----RREERASVDYNQDRTYESVRTPTYPE 644
 QY 1081 -STDDKHLKQPHFGQBCVSP--YASRGANGSENVGSNGINQNVQSQCQDDYED 1137

DB 645 DSRRDYPARGREFYSEWETVQGYESRYDDPREVDYRNDPEQDI-----REYSYRQ 699
 QY 1138 DKPTNYERYSEEQHEEERPTNYSIKYNEKRHVDPID--YSLKYATDIPS-SQKQS 1194
 DB 700 RERERERFESDRDRDHERRP---IERSQSPVHLRPOSPGASPSQABRLPSDSERRL 755
 QY 1195 FSPKSSSGGSSK-----TEHSSSSENTSTPSSNAKQCNQHLHSSAQ 1237
 DB 756 YSRSDRSGCSSLSPRYEKLDKRLERYTKNEKTDKERTFDPERVERERIRKEKVE 815
 QY 1238 --SRSGQPKAAATCKVSSINQETIQTVCVDDTTCFSRCSLSSLSAABEIGCNOTTQE 1295
 DB 816 KDKTDKQKRKGVHSPSSQSETDQENEREQSPKPRSCNKLSEKADKEGIAKNR--- 871
 QY 1296 ADSANTLOIAIEKIKIGTRSAEDPVSEVPVAVSOHPRTKSSRLQ-----GSLSS 1344
 DB 872 LELMPCVVLTRVKEKGV-DHTPVEKAKLNDTVKSSALDQKLQVSGTEPAKSDLSK 931
 QY 1345 -ESARHKAIVE---PSSGAKSPKSGAOTPKSPPEHYVQETPLMFSRCTSVSSLSDSPESRS 1400
 DB 932 LESVRMKVPKEKGLSSHVVVEKEGLKAR--KHLKPEQADGVANVLEKLEAKGRF 988
 QY 1401 IASSVQSEPCSGMVSGIISPSDLPSDQTMPPSR---SKTTPPPQTAQTKREVFNKA 1457
 DB 989 ADSNLKAEK-----QKPEVKKSP--EMEDARVLSKKQPDVSSREVILLREGEAER 1038
 QY 1458 PTAB---KRESGPKQAAVNAVQVLPDADTLHHPATESTPDGFCSSSLGALSLEP 1514
 DB 1039 PVKEILKRES-----KKIK-----LDRINTVASPKDQELASISVGGSGRP 1080
 QY 1515 FIQDVELRIMPVQENDNGNTESEOPKESNENQEKAEKTDIDSEKDL-----LDDSD 1569
 DB 1081 --SSDLQARL-----GELAGESVENQEVQSKKPIPSKPOLKQLQVLDDQGP 1124
 QY 1570 DDTEILEE-CII--SAMPTKSSRKAKKPAOTASKLPPPVARKPSQLPVYKLLSQRLQP 1626
 DB 1125 EREDVRKNYCSLDETPERKSGQKSHSVNTEKIGIDHTQS-----YKQMEQRRKQ 1180
 QY 1627 QKHVSF-----TPGDDMPRVYCVBETPINFSTATSLDLTIESPPNELAAGEV--- 1675
 DB 1181 QMEIEIAKSEKFGSPKQVDE--YERSLVHEVGKPPQDVTDSDPPSKKRMHDVDFDI 1237
 QY 1676 -----RGAQSGEPEKDDTIPTGRST-----DEAQCKTSSVTI---PBLDDN 1716
 DB 1238 CTKEERNYRSRQISEDSER---TGGSPSVRHGSHFEDDPICSPRLLSVGKSPKVDK 1293
 QY 1717 KAEBGDILASINSAMPKSKHFKPFVKIMQVQ-QASASSAPNKNQLDGKKK--- 1771
 DB 1294 VLPYSNI---TVREESLKFPYDSSREEQVADWAKIKLVNSEDDELNRWDSQMKQDAGR 1350
 QY 1772 ---PTSPVKPIQNTYETVRKNADSKNNNAERVSFSDNKKDKKKONKJNNS----- 1820
 DB 1351 FVSPFNSIHK-----RDSLKR--SVRDLEPGEVPSDSEDEGEH--KSHSPRASAL 1398
 QY 1821 -----KDFNDKLPNNEDVRG-----SFAFDSPHVTPTEGTPYCFSDNSL 1862
 DB 1399 YESRLSFLDRDEDKLRERDERLSSLERNKYFSALDK---TTPDTKALLERAKSL 1454
 QY 1863 SS-----LDFDDDDVDL-----SREKAE-----LRKAKENKSEAKV---TS 1896
 DB 1455 SSSREENWSFLWDSRFANFRNNKDEKVDSPAPRIPSWYMKKKKIRTDSGKDDKED 1514
 QY 1897 HTELTSNQOSANKTQ---AIKQPINRGOPKPILOKQSTFPQS---SKDIPDRGAATDE 1949
 DB 1515 HKEREQKQLFASRLPHSHSIFEDSKRLOHLERKEEDSDFISGRYKQTSKANSTTD 1574
 QY 1950 KLOQFAIENTFVCFSHSNSSLSLSDIDQENNNKENEPKETEPPDQSGPFSQASGYAP 2009
 DB 1575 SIQE-----PVVLFH--SRPELTMQKQEKQKQKPEKQEDTENHPKTPES---AP 1624
 QY 2010 KSHVBE-DTPVCFSRNSSLSSLSIDSDLLQBCISSAMPKKKPKSRLLKQDNKHSPRNM 2068
 DB 1625 ENKDSLUKTPPSVG--PPSVTVTLESAPSALE-----KTTGDKTVEAP--- 1666

QY 2069 GGIIGEDLTLDLKDQIOPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAACLSRQA 2128
DB : : : : :
QY 1667 --LVTEETVE-----PATVSEAPAPAPVQEQVD-----LPPGA 1706
DB : : : : :
QY 2129 SSDSDIISLXSGISLSPGPHLTDPQEEKPTSNKGPRLKPKGKSTLETKIESKGI 2188
DB : : : : :
QY 1707 DPDXEAM-MPAGVEEGSGGQPPYLDKAPT-----PGASFQASNVDP 1752
DB : : : : :
QY 2199 KGGKKVYKSLITGKVRNSEISGQMKQPLQANMPSISGRGRWMIHIGVRNSSSTSVSK 2248
DB : : : : :
QY 1753 -----PDSQPLSK 1761
DB : : : : :
QY 2249 KGPPPLKTPASKSPSEGOTATTPRGAKEPSVKSELSPVARQTSQ----- 2291
DB : : : : :
QY 1762 -----PAQSEBANE-----PAAEKDADADAEADQVAAAPESQPPASEDLEVP 1809
DB : : : : :
QY 2292 -IGSSKAPSGSGRSDTPSPRA-----QQPLSRPIQSPGRNSISPGRNGISPPNKLSQL 2345
DB : : : : :
QY 1810 PVAADKKPNKS-KRSKTPVQAAAVSIVEKPVTRKSERIDREKLKNSPRGEAQKLEL 1868
DB : : : : :
QY 2346 PRTSSPSTASTKSSGSGWSTSPGRQMSQONLTKQGLSKNASSIPRSESASKGLQNM 2405
DB : : : : :
QY 1869 KWEAEKITRTAKNSAADLEHPELSPLSR-----TERRNV-----RSVATMGDHE- 1915
DB : : : : :
QY 2406 NGGANKKVELSRMSTKSSGSGSDRSRSPVLRQSTFIKEAPSPTLRRKLEBSAS- 2461
DB : : : : :
QY 1916 NRSVPKPEVQPRVTRKELRELQEAAPVTPPRG-----RPPKTRRADDEBENEAKE 1970
DB : : : : :
QY 2462 -FESLSPSRASP-----TRSQATPVLSPL- 2488
DB : : : : :
QY 1971 PAETLKPEGRNRSRQKTAAGGPGQKKGKNEPKVDATREATEV- GPOIGVKESME 2029
DB : : : : :
QY 2489 PMSLSTHSSVOA-----GGWRKLPPNLSPTIEVNDGRPA--KRHDIAHSHSESPSLPI 2541
DB : : : : :
QY 2030 PKAABEEAGSQKDRKDAGTKNPPETAP-VEVVEKKPAPEKNSKKGESRY-SRLAV 2087
DB : : : : :
QY 2542 NRSQTKWEKHKHSSSLPRVSTWARTGSSSLASGESSKAKSED-----EKHVNISG 2597
DB : : : : :
QY 2088 DKSASLKNVDAVS---PRGAAOAGERESGVAWSPEKSPQKEDGLSSQLKSDVPD 2144
DB : : : : :
QY 2598 TKQKENOVSAKGTWRKTKENEFSTNSTSTVSSGATNGAESKTLVQ-----MAP-- 2649
DB : : : : :
QY 2145 DKEPEKDVASGP--SPEATQAKQMELEQAVEHIAKLAFAASAAYKADAPGLAPED 2202
DB : : : : :
QY 2650 -----AVSKTE-----DYWVRIEDCPINNRSRSGSFTGTPPV----- 2682
DB : : : : :
QY 2203 RDKPAHQASSETLAAAGISLINDISGERENFPAPPVPEGESQTDLPAGAAQALQPSSEG 2262
DB : : : : :
QY 2693 -----IDSVEKANPNK-----DSKDNQAKQVNGSVPMRTVGLNLR 2721
DB : : : : :
QY 2263 METDEAVSGILETAATESRPPVNPADPPSAGPTDTKEARNSSETSHSVP-EAKGSKEV 2321
DB : : : : :
QY 2722 LNSFTQVDAPPQKGTETKPGQN-----NPPVNSE-----TNESIVE--RTP 2761
DB : : : : :
QY 2322 EVTLVRKDKGQKTRSRKENTKVVAVPESHVPSNQAGSPAAEGTTVQHSAP 2381
DB : : : : :
QY 2762 FSSSSSSK-HSPSGTVAARVT--PNNVNS-----PRKSSABSTARSPPQITPPV 2809
DB : : : : :
QY 2382 QEEKQSEKPHSTPQSCSTSDLSKIPSTENSSQIESVEERTKASVPPDLPPPPQ-PAPV 2440
DB : : : : :
QY 2810 NNNTKRDSDTDSSESSGTSQP 2831
DB : : : : :
QY 2441 DEEPOARFVHSIETSDPVTTP 2462
DB : : : : :

RESULT 7

ID MINT MOUSE
AC Q62504; Q30TN9; Q99PS4; Q9QZM2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC TISSUE=Testis;
RX MEDLINE=99379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mxx2 binding protein, recognizes and
RL regulates the rat osteocalcin promoter.";
RN Biochemistry 38:10678-10690(1999).
RP SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA Ono M., Ohnato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watahiki A., Muramatsu M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT J/SH activity.";
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 318-578 FROM N.A.
RC TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozat F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fzanes C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I mvosins.";
RN Genomics 40:332-341(1997).
RP SEQUENCE OF 2598-3644 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RN DNA Res. 10:35-48(2003).
RP TISSUE SPECIFICITY
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kosczka U., Astrahantseff K., Bourteelle S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RL pathway.";
RN EMBL J. 21:5417-5426(2002).
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=22483652; PubMed=12594956;
RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
RT of Notch/RBP-J signaling pathway.";
RN Immunity 18:301-312(2003).
CC -1- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative

Db 2571 -----PPCDPKH-----PPLEGVAAAAPN-----ADTQASE----- 2597

Qy 2720 NRLNSFIQVADPOK-GTEIKPG-----QNNPVVSETNESSIVERTPSSSSSSKHSPP 2773

Db 2598 -----VFAADKEKAVPIAPKITSVIRMPVSIDLENSQKITLAKAPQTUTGLVSAL 2651

Qy 2774 SGTVAARVTPNTNPSPRKSSADSTSRPQIETPVN 2810

Db 2652 TGLNVSLVFNALKGVPKGVATLKLGLVSTPAGPVN 2688

RESULT 8

ID_ZEP1_MOUSE STANDARD; PRT; 2688 AA.

AC Q03172,1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Zinc finger protein 40 (Transcription factor alphaA-CRYBP1) (Alpha A-

DE crystallin-binding protein I) (Alpha A-CRYBP1).

GN HIVEP1 OR ZNF40 OR CRYABP1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=DRA/2J;

RA MEDLINE=95138112; PubMed=7836383;

RA Brady J.P., Kantorow M., Sax C.M., Donovan D.M., Hamada K., Sax C.M., Norman B.,

RA Flanagan J.R., Ozato K., Westphal H., Piatigorsky J.;

RT "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA

RT encoding a protein that binds to a cis sequence motif shared with the

RT major histocompatibility complex class I gene and other genes.";

RL Mol. Cell. Biol. 10:3700-3708(1990).

RN (2)

RP SEQUENCE OF 2024-2688 FROM N.A.

RC TISSUE=Lens epithelium;

RX MEDLINE=90287161; PubMed=1694016;

RA Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B.,

RA Flanagan J.R., Ozato K., Westphal H., Piatigorsky J.;

RT "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA

RT encoding a protein that binds to a cis sequence motif shared with the

RT major histocompatibility complex class I gene and other genes.";

RL Mol. Cell. Biol. 10:3700-3708(1990).

CC -1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE

CC PALINDROMIC SEQUENCE 5'-GGGAATCCC-3' IN THE ALPHA-A CRYSTALLIN

CC PROMOTER.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY

CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH

CC ZINC-FINGER IN-BETWEEN.

CC -1- SIMILARITY: STRONG, TO HIVEP2.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L36829; AAA98810.1; JOINED.

DR EMBL; L36825; AAA98810.1; JOINED.

DR EMBL; L36822; AAA98810.1; JOINED.

DR EMBL; L36827; AAA98810.1; JOINED.

DR EMBL; L36828; AAA98810.1; JOINED.

DR EMBL; X68946; CAA48762.1; JOINED.

DR PIR; I49477; I49477.

DR HSSP; P15822; 32NF.

DR TRANSFAC; T00007; -.

DR MGD; MGI:96100; HIVEP1.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 5.

DR SMART; SMO0355; Znf_C2H2; 5.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.

KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 165 170 POLY-SER.

FT ZN_FING 407 429 C2H2-TYPE.

FT ZN_FING 435 459 C2H2-TYPE.

FT DOMAIN 799 802 POLY-SER.

FT DOMAIN 851 854 POLY-PRO.

FT ZN_FING 953 976 C2HC-TYPE (POTENTIAL).

FT DOMAIN 1482 1486 POLY-SER.

FT ZN_FING 2074 2096 C2H2-TYPE.

FT ZN_FING 2102 2126 C2H2-TYPE.

FT DOMAIN 2179 2182 POLY-ASP.

SQ SEQUENCE 2688 AA; 288341 MW; 5EAD46C3A7008BE6 CRC64;

Query Match 2.8%; Score 404; DB 1; Length 2688;

Best Local Similarity 18.4%; Pred. No. 1.3e-07;

Matches 463; Conservative 328; Mismatches 896; Indels 826; Gaps 106;

Qy 1007 KIHSANHMDNDGELDTPINYSKYSDQLNSGRQSPSQNRWARPKHII-HDIKQSEQ 1065

Db 6 QIHPRN-LRDKIEBAQKELNGAEVSKVELEAGVGTSESLKGVKRVKIVAEHLEKIPK 64

Qy 1066 RQSRN-----QSTTVPVVT--ESTDDKHLKFPQHFQQECVSPYSRGANGSETNRVG 1116

Db 65 SPLNPLQTHKQNTPEPPFVLSASESHVK-----HNCVPAKQGR-----QFTKQNG 113

Qy 1117 SNHGINQVQSLOCOEDDYEDDKPTNYSERYSE-EOHEEERPTNYS-----IKYNE 1168

Db 114 ETPGTAESSES-----GDLVSPKKTSSPHQSRSELRWRSESDPTRLSLDGLDQGDSSSS 169

Qy 1169 EKRHVDPQIDYSLXYATDIPS-----SOKQSFSPKSSSGSQSKT 1208

Db 170 SKARTDNSECSPPCSTTPPSYTTAFDVLKAMEPELSTLSQKSSCAIKTEKLRNKT 229

Qy 1209 EHMSSSENTTPSSNAKRNQL-----HFSSAQSRSGQPOKAATC----- 1249

Db 230 VRSPSKLNSLDAPNATSDPLVVEPCPPCTSYFHVASTQSEQVAAQCVSHLYSSQD 289

Qy 1250 ----KVSSINQE-----TIQTYCVEDT---PICFRCSSLSLSAEAEIGCNOT 1292

Db 290 HLVPKLSQQNQOLPGLHGTGSLNLTLESTKLEPIYNTAVTSTVGLTSPSTRTQVTPP 349

Qy 1293 TQEADSANTLQIAIEKIGITRSAEDPVSE---VPAVSQHP-----RTSSRLQGS 1340

Db 350 HQQMDVSPLSVSPAS---STQSPGPIYSSAHVASVVSQVEQMCSSLRLDQKFKKQK 406

Qy 1341 -----SLSESARHKAVERFSSGAKS-----PSKSGAOTPKSPPEHV----- 1377

Db 407 YICEYCNACAKPSVLLKHRSHTGERPPCVTCGPFKTKSNLYKHKSHAHTIKGLV 466

Qy 1378 -----QETPLMFRCSTSVSLDSPESRIASSVOSEPC-----SGMVSGIISP 1420

Db 467 LQPEAGGLFSQECFKALSVDIEDSGEDGLADGRQNNPCVKDLQVPQTKVTSNP 526

Qy 1421 SLDLPSDQGNTP-----PSRSKTPPPPPQTAQTKRVPKRNKAPTAKRES 1465

Db 527 ESLP-----KLIPSNSDHVRGFSQSDRPSQAPTELKPVVHVPMPLKTKDCLQVANP 582

Qy 1466 GPKQAANVAQRVQVLPDADTLTHFATESTPDGFCSSLSLSALSDLEPIQK-DVELRI 1524

Db 593 NPELPS-----PQSPRDLHVASI-----LSHASVSSLEMDSECHQGDVIOSE 626

Qy 1525 MPVQENDNGNETESQPKSNEQKAEKTDISEKDL--LDDSDDDDIIELEECI--- 1579

Db 627 GKP--DSHSGTAHAQLQRQATQEDPQOQKLLSPSLGSTDGYSFRRSEADQAVSP 684

Qy 1580 -----ISAMPTKSRKAKKPAQTASKLPPPVARPSQLPVY 1615

Db 685 TPFARTPTTDDPDKNGGAPGPRISAPASALATGKSSVVTGQMRPPLATKLEERIS 744

Db 733 SYIQD-----ETIPGSETEQITISDEIHDPEERPAAPRPHSTYDLPGEAGAPFEAS 787
 Qy 1038 -----SGRQSPQNERWARPKHLEIEIKQSEQRQSNOSTYVYVTESTDDKHL 1087
 Db 788 QPADSAVATSKVXGTETELTYTNIAAPLABEEHVSAITSEC-----DKLS 839
 Qy 1088 KQPFGQOECV-----SPYRSGANGS-----ETNRVGSNHGINQ-----NVQSLS 1129
 Db 840 SFATSAVEDQSASLTAPOTEBTGKSSLLDVTISIPSSRTAQTGLDYVPSAGTISPTS 899
 Qy 1130 COEDD-----YEDDKNTYSER-----YSEEEQHEEERPTNYSIK----- 1165
 Db 900 SLEEDGKSPCEDFSVTGSEKKEGIIIGKLSGERAVEEBEETANVEMSEKLCQYQ 959
 Qy 1166 -----YNEKKRHVDQPIDYSLKYATDIPS-----S 1190
 Db 960 TPVFSAPGHALPGEPALGEAEERCLSPDDSTVKVASPPSPGPSATHPTPHQSPVEKS 1019
 Qy 1191 QKQSFSPKSSGQSKSTEHM-----SSSENTSTP-----SSNAKROQLHPSAQSAGQPOK 1245
 Db 1020 EPQDFQ-EADSGWDTKRTPGVGKEDAAEETVKPGPEEGTLEKEEKVPP-----PRSPQAE 1074
 Qy 1246 AATCKVSSINQIOTYCVEDPTICFSRCS-----LSSLSAEDDEIGCNOT 1292
 Db 1075 APVNIDEGLTGCTIQLLPAQOKXAIYFEIMEAGEPTGPILGAELPGGLRTILPQFGRPOK 1134
 Qy 1293 TO-----EADSANTLQIAIEIKETGTRSAEDPVSEVPAVSQH-----PRTKSRLL 1337
 Db 1135 DEVLRYPDRSLSPEDAESLSVSPSPDTANQEPFKSPCGLTQOYLHKDRWPVSPEDT 1194
 Qy 1338 QGSSLSSES-ARKHKAVERSSGAKSPKSGA-----QTEPKS 1371
 Db 1195 QLSLSSESPSKETSLDVSQKQSPESIGTLQFGLNLGKEMGHLMQAEOTSHHTAPMS 1284
 Qy 1372 PPE-HYVOETP-----LMFSRCTSVSSLDSPESRSIASVQSEPCSG-----MVGSIISPS 1421
 Db 1255 VPEHAATASPTDGTTRYSAQTDITD-DLDRKSPASSFSHSTSPSGNGKYLPGAITSPTD 1313
 Qy 1422 DLPDSFGQTMPSRSTPPPPQTAQTKRE--VPKNKAPTAE-KRESGPKQAAV-----NAA 1475
 Db 1314 EHILTPDSFSKSPESLPQPALEDIAIKWEDKVPGLKORTSEQKEPEPFQEVLOQKDKT 1373
 Qy 1476 VORVOVLPAADT-----LLHFATESTPD-----GFCSSLSAL-SLDEPFIOKD 1519
 Db 1374 LSHKEVVEPKDAILYKQBALHVKEAVKQOKALEQKGRDLEQDKTALQKDALEPKD 1433
 Qy 1520 VELRIMPPVOENDNGNETSEOPKESNENQEKYAEKTTIDSEKLLDDSDDDIE-----IL 1575
 Db 1434 KDL-----EEKDKALEQKDKIPEE--KDKALEQKDTALEQKDALEPKDKLEQKORVL 1485
 Qy 1576 BECIISAMPTKSSR-----KAKKPAQTASKLPPPVVARPKSOLPVVYKLLPSQNRLOP 1626
 Db 1486 EQ--KEKLEPKDKALDQKVRSEVHEKAPEDTVAEMKDRDLEQTKAPEQK-----HQAE 1538
 Qy 1627 QK-HVSFTPGDDMPRYVCVETGPIINFSTATSLDITIESPPNELAAGEVGRGAQSGBE 1685
 Db 1539 QKDKVSEKKDQALEQKYWALG-----QKDEALEQNIQALEENHQOT 1579
 Qy 1686 KXDITPTEGRSDTEAGCGKTSVTIPELDDNKABEGDILAE-----CINSAMPKG 1735
 Db 1580 EQESLVQE-----DKTKPKKOLBEKPEKVMKEEKLEALLETKALGULEESLVQEGRAE 1635
 Qy 1736 KSHKPFVKKIMDQVOQASASSAPKNQLDGKKK-----KPTSFVKPIPQNTY----- 1786
 Db 1636 QEEKYWRGQDVVQEWQETS-----PTREEPAGEQKELAPAWEDTSP-----EQDNKYWRGR 1686
 Qy 1787 TVRVKNADSK-NNLNARVFP-----SDNKDSKQNLKNNKSKDFNDKLPNNEDRV----- 1834
 Db 1687 EDVALEQDITYWRELSCEKRWVFPFHELDGQGARPHYTEERESTFLDEGPDDEQEVPLREHA 1746
 Qy 1835 -RGSPAFDPSPHVTPTEGTPYCFSRNDSLSLDFDDDD-----VLSREKAEKLRKAKENKE 1889
 Db 1747 TRSPWASDFK-----DFQESSPKGLEVERMLAESFVGLPPEE 1784

Qy 1890 SEAKVTSHTELTNSQOSANKTOAIAKQPINRGQPKILOKQSTFFQSSDKPIDRGAATDE 1949
 Db 1785 EDKLTSPPEIILSPAS-----PPENVGQRPSPAPQESFIP-----DP 1823
 Qy 1950 KLONFAIENTPVCFSHNSLSLSLSDIDQNNKNEPIKETEPPDQSGPSPKQASGYAP 2009
 Db 1824 KLMPH-KMNEP-----TTPSWLADIPW--VPKDRPLPAPLSPAPGPPPTP-----AP 1868
 Qy 2010 KSFHVEDTPVCFGRNSSLSSLSIDSEDDLQECISSA-----MPKKKPSRLKGDNEKHS 2064
 Db 1869 ES-----HTAPFWSGCTP-----EYDSVVAVQEGAAELGGPSPGLKYKVGXERE-- 1919
 Qy 2065 PRNMGILGSDTLDLKDIQRPDSEHGLS-----PDSENPDKWIAIOEGAN 2109
 Db 1920 -----GRAEAPDKSHXSKVPKARKSHATTEPEQTEPEOREPTPYPDERSFOVADIYE-- 1972
 Qy 2110 SIYSSLHQAAAAACLSRQ-----ASSDSDSILSKSIGLSGSPHPLTPDQEEKFTNKGP 2165
 Db 1973 ---QMLMTGLGPACPIREPPPLGAAGWPPCLSTKEAAGR--TSAEKELSPISP 2023
 Qy 2166 RILKPEKSTLETKKJIESKGIKGGKVYKSLITGKVRNSNISEIGQMKQKLOANMPSIS 2225
 Db 2024 ---KSLQSDPTF-----SYAALAGPTVPXPXEPGSPME 2054
 Qy 2226 RGRTHIHPGVRNSSSSTSPVSKGPKLTPASKSPSEGTATTSPRGAKPSVK----- 2279
 Db 2055 PSUTPPAVP-----PRADILSKGP-----SPPLNGNILSCSPDRRSPSPKESGRSH 2100
 Qy 2280 -----SELSPVARQTSQIGGSKAPSRSGSRDSTSPPAQOFLSRP-----IQS 2323
 Db 2101 WDDSTDSSELEKAREQAE-----KEAQSPSPHPHPIMSGSPPLPWPTEAHVSP 2148
 Qy 2324 PGRNSISPGRNGISPP-----NKLQPLRTSPSTASTKSSSGSGKMSYT-----SPG 2370
 Db 2149 PLXSHLGXARPSLDFPASAFGSLEXP-POXPSPAEPRSPAPCGSLAFSGDRALALAPG 2207
 Qy 2371 ---RQMSQOQLTKQGLSKNASSIPRSES-ASKGLNOMNGNGANKVKELSRMSTKS 2424
 Db 2208 PPRTRRYDELEVTKAPSLD---SSLPQLPSPSPGXPLSN-----LPRPASPAL 2255
 Qy 2425 SGSESRSERPVLRQSTFTIKEAPSPTRLRKLEESAFE-----SLSPSSR 2470
 Db 2256 SEGSSSEATTPTVI---SSVAERFSPSLEAAEQESGELDPGMEPAAHXLWDLTFLSPAP- 2310
 Qy 2471 PASPTRSQATPVLSPLPDMSLSTHSSVQAGWRKLPPLNLSPTIEY-----NDG 2520
 Db 2311 PASLDLALAPSLPGDMGDIILPCHLECEAATEKPSPOVFSQVSEDAANGPTTETSNNPP 2370
 Qy 2521 RPAKRHDIAHSHGESPSRLFINRSWTW---KREHSHKSSSLPRVSTWRTTSSSSSILSA 2576
 Db 2371 XPAP---AKAENEEAAAAXPAWERGAWPEGAERSRSDTXXLSPEQVPCVXGGSGPPSSA 2426
 Qy 2577 SSE---SSEKAKSEDEKHVNSISGT-----KQSKENQVSAKGTWRLK--KE 2617
 Db 2427 SPEVEAGPQOCXTEPRPHRGELSPFLNPLPPSIDDRDLSTBEVLVGRGRRRVGGPG 2486
 Qy 2618 NEPSPTNSTGTSVSSGATNGAESKLIYOMAPAVSKTDETVVWRIEDCP----- 2665
 Db 2487 TTGXFXPVTDPTPTSASDSSSQS-----DSDVPPEETEECPSTAEALDSDSE 2535
 Qy 2666 -----INNPRSGRSPGTNTPTPVIDSVEKANPNKIKSDKNQAKONVNGS 2710
 Db 2536 DGDFLPVDXKXGVSGTHHPRPGHDP---PPL-----FPDXRPS----- 2571
 Qy 2711 VPMTVGLNRLNLSFIOVDAPDQKGTBIKPGQNNPVVPSVETNESSIVERTPFSSSSSKH 2770
 Db 2572 -PPAP-----DVCMADPEGLSESGRVERLEKXKVKQGRVORRAP-----GKDKP 2615
 Qy 2771 SSPSGTVAARVTPFNYPNPSPRKSADSTARPQIPTPVNNNTKKRDKSTDTSESSG 2827
 Db 2616 VSPXRRLKLR---QKRGPTGKGSXDRVSRXPKRSRSTXQVTPABEKGQHSFMSKG 2669

QY	897	IHTSQEDRSGSTTELHCVTDBENALRRSAAHTSNTYNTFKSE--NSNRCTCSMPYAKL	954
Db	1421	IYTKSESQOEQEBEID-MTSEKN-----DETETSTSVLKSHLVNEVPVLASPDLLS	1472
QY	955	EYKRSSND--SLNSVSSSDGYGKRG--OMKPSIESYSEDDSEKFCSYGOYPADLAHKHS	1010
Db	1473	EVSEMAQDLIKMTAILTTDVSDKAGSIKVKELVKAABEEBPGEFF-----	1516
QY	1011	ANHMDNDGELDTPINYSIKYSDEQLNSG-----RQSPSQ-----NERWAPKHIIE	1057
Db	1517	-----EIVERVKEDLEKVEILIRSGTCTRDESSVQSSRSERGLVSEEW-----IYS	1563
QY	1058	DEIKQEQRQSRNQSTTYPVVYTESTDCKHLKFQHPFGQOECVSPYRSRGANGSETNRVGS	1117
Db	1564	DEBIEARQKAPLEITEYPC-VEVRIDKEIK-----GKVE-----KDSGLVNVLTDDLNT	1613
QY	1118	NGHINQVQSGLCQEDDYEDDKTNYSE-----RYSEEQHBEERPTNYSIKYNEKHH	1172
Db	1614	CVPLPKEQLQTV-----QDKAGKCEALAVGRSEKEGKOJPPDETOSTQKHPSLGS	1666
QY	1173	VDOPIDYSLYKATDIPSSOKQSPFSKSSSGQSKTEHMSSSSENTSTPSSNAKRONQLH	1232
Db	1667	IKAPVRKULK-----EKQVQ-----KEGLQASAKA-----ELK	1696
QY	1233	PSSAQSRSGQ-----PQAACTKVSSINQETIQTVCVEDTPICFRCSLSLSSSAEDRI	1287
Db	1697	KGSSEESLGEDPGLAPEPLTPVKATS-----PLIETPTI-----GSIKDKV	1737
QY	1288	GCNQTQEADSANTLQIAIEKEKIGYRSADPYSEVPANSQH-----PRTKSSRLQGS	1340
Db	1738	KALQKRVDES-----QKGRSKLPIRVKGBED-----VPKKTTHRPHPAASPLKSERHAP	1788
QY	1341	SLSSSEARHKAHVFFSGAK-----SP-SKSGAQTPKGPPEHYVQETPLMFRCSTSS	1389
Db	1789	SPSPKTERHSTL--SSSAKTERHPVPSPSKTEKHSVPSSAKTERHSPASSSSKTEKHS	1846
QY	1390	VSSLDSFESRSSTASSVQSE---PCSGWVGIIISPDLDPSQCTWMPGRS-KTPPPPPQ	1444
Db	1847	PVSPSTKTERHSPVSTKTHERHP-----VSPSGKTDKRPPVPSGTEKHPVSP--	1897
QY	1445	TAOTKREVPKNKAPTAEKRSGEKQAAVNAVQVLPDADTLHFATETSTPDGFCSS	1504
Db	1898	-GTEKRLPVSPGRDTHQ-----PVSTAGTEKHLVSPSS--GKTEKQP-----	1940
QY	1505	SLSALSDBEPIQKDVLR--IMPPVQ--ENDNGNETESEQKESNENQEKAEKT-IDSE	1560
Db	1941	PVSPSTKTER-IETWVSRELMKAFQSGQDPFKHKTGLFHKHAKQKQPOEKGKRVKE	1999
QY	1561	K-DLLDSDDDDDJELLEECIIISAMPKSSRKAKKPAQTAS-----KL	1601
Db	2000	KGPILTQREAQKTENQTIKQGBQLPVTGTAESKRGVRVSSIGVKEDAAGKKEKVLSHKI	2059
QY	1602	PPPVARKP-----SOLPYKLLPSQN-----RLQPKHVS-----FTFGDDMPR	1640
Db	2060	PEPVQSVPEESHRESEVPKEKMADEQMDLQISPRKSTSTDFSEVIKQLEDNDKYQQ	2119
QY	1641	VYCVEGTPIFNSTATSLSDUTIESPYY-----ELAAGSVRGGAQSFGFEKRDIT	1690
Db	2120	FRLSEET-----EKAQHLQDLVLTSPFNFTPLDYMKDEFLPALUSGALDGSE---	2172
QY	1691	PTGRSTDEAQGGKTSVTT--IPELDNDKAEGBGIIAELACINSAMPKSHKHPRVKIMD	1748
Db	2173	KNEG-----VAGSPCGSIMEGTFOISEESYKHEGLAETETS-PESLSFSP---KXSEE	2223
QY	1749	QVOQASASSAPNNKQLDGKKKXPTSPVKPIPONTEYR--TRVRKNADSKNINLAERVFS	1806
Db	2224	QTGETKESTKTETTTTIRSEKEHPTT--KXITGGSERGGATVTEDSSTSTESFQKEATLG	2281
QY	1807	DNKDSKKQNLKNNSKDFNDKLPNNEDVRVGS-----FAFDSPHYVTPLEGFPYCFSRNDSL	1862
Db	2282	SPKDOTS-----PRQDDCTGSCSVALAKETP-----TGUTEBAADEGQRTF	2323
QY	1863	SSLDF-----DDDDVDLSREKALRKAKENKESAKVTSHTLSTNQOSANKTOIAKQPIN	1919

Db	2324	GS	AHKTQ	DS	SAQ	S	T	S	D	E	T	K	A	L	P	E	A	S	V	K	T	D	T	G	T	E	-----	S	K	P	G	V	I	R	S	P	Q	2375																		
QY	1920	RQ	CP	I	L	O	K	Q	S	T	F	P	Q	S	S	K	D	I	P	D	R	G	A	T	E	K	I	O	N	F	A	E	N	T	P	V	C	F	S	H	N	S	L	S	S	D	I	S	D	I	Q	E	N	1979		
Db	2376	L	E	L	A	L	P	-----	S	R	D	S	E	V	L	S	A	V	A	D	D	S	L	-----	A	V	S	H	K	D	S	I	E	A	S	P	V	I	E	D	N	S	2414													
QY	1980	N	N	K	-----	E	N	E	P	I	K	E	T	E	P	D	S	-----	O	G	E	P	S	K	P	O	-----	A	S	G	V	A	P	K	S	F	H	V	E	D	T	P	V	C	F	S	R	N	S	L	S	S	L	S	I	2032
Db	2415	S	H	T	P	D	S	L	E	P	S	K	E	P	C	R	D	S	L	E	S	P	V	E	P	K	A	G	I	P	P	S	H	F	L	-----	P	A	A	V	A	K	T	-----	2463											
QY	2033	D	S	E	D	I	L	L	O	C	I	S	A	M	P	K	K	K	P	R	L	K	G	D	N	E	K	H	S	P	R	N	G	G	I	L	G	E	D	L	T	L	D	L	K	D	I	O	R	P	D	S	E	H	L	2092
Db	2464	-----	E	L	L	T	E	V	A	S	V	R	-----	S	R	L	L	R	D	P	D	G	S	A	-----	E	D	S	L	E	O	T	S	L	M	E	S	S	G	K	S	P	L	2503												
QY	2093	S	P	D	-----	S	E	N	F	D	K	A	I	O	E	G	A	N	S	I	V	S	L	H	O	A	A	A	A	C	L	S	R	O	A	S	S	D	S	D	S	L	S	L	K	S	G	I	S	L	G	S	P	F	H	2149
Db	2504	S	D	T	P	S	S	E	V	S	E	V	T	P	K	T	T	-----	V	S	T	P	K	P	A	V	I	H	E	C	A	E	D	D	S	E	N	-----	2544																	
QY	2150	L	T	P	D	O	B	E	K	F	T	N	K	G	P	R	I	L	K	E	K	S	T	L	E	N	-----	K	I	E	S	E	K	I	G	K	K	V	I	S	L	I	T	G	K	V	R	2204								
Db	2545	-----	E	K	K	F	T	P	E	-----	E	M	F	Q	W	T	K	I	M	F	D	E	L	E	O	E	A	K	-----	Q	K	R	D	Y	K	-----	2579																			
QY	2205	S	N	S	E	I	S	G	Q	M	K	P	L	O	A	N	M	P	S	I	R	G	R	T	I	H	I	P	G	V	R	N	S	S	S	T	P	S	V	K	G	P	P	L	K	T	P	-----	A	S	K	S	P	2361		
Db	2580	-----	K	E	P	K	Q	-----	E	S	S	S	D	P	A	D	C	S	V	D	V	E	P	K	H	T	G	S	G	E	D	-----	2612																							
QY	2262	S	E	G	O	T	A	T	T	P	R	G	A	K	P	V	K	S	E	L	S	P	V	A	R	O	T	S	O	I	G	S	S	K	A	P	S	R	S	G	R	D	T	P	R	P	A	O	C	P	L	S	R	P	I	2321
Db	2613	E	S	G	V	P	V	I	T	S	E	R	K	V	S	S	E	S	E	P	E	L	A	Q	L	K	G	A	D	S	G	L	P	-----	E	P	V	I	R	-----	2655															
QY	2322	Q	B	F	R	N	S	I	S	P	G	R	N	I	S	P	P	N	K	L	-----	S	O	L	P	R	T	S	P	T	A	T	K	S	G	S	G	M	S	Y	-----	T	S	P	G	R	Q	M	2373							

```

3195 VQTGDIPPLSG--VKQISCPDSSE 3216
DB
RESULT 11
PCGV CHICK STANDARD; PRT; 3562 AA.
AC Q09053; Q09045;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
GN Chondroitin sulfate proteoglycan core protein 2 (PG-M).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS V0 AND V1).
RC STRAIN=White leghorn; TISSUE=Limb bud;
RX MEDLINE=93300845; PubMed=8314802;
RA Shinomura T., Nishida Y., Ito K., Kimata K.;
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multiforms of PG-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q09053-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q09053-2; Sequence=VSP_003093;
CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X60226; CAA42787.1; -;
DR EMBL; D13542; BAA02742.1; -;
DR PIR; A47171; A47171.
DR HSP; P00740; 1EDM.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR Pfam; PF0000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.

```

```

DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 3562
FT DOMAIN 27 143
FT DOMAIN 166 243
FT DOMAIN 264 345
FT DOMAIN 3254 3290
FT DOMAIN 3292 3328
FT DOMAIN 3341 3455
FT DOMAIN 3460 3518
FT DISULFID 44 129
FT DISULFID 171 242
FT DISULFID 195 216
FT DISULFID 269 344
FT DISULFID 293 314
FT DISULFID 3269 3269
FT DISULFID 3263 3278
FT DISULFID 3280 3289
FT DISULFID 3296 3307
FT DISULFID 3301 3316
FT DISULFID 3318 3327
FT DISULFID 3334 3345
FT DISULFID 3362 3454
FT DISULFID 3430 3446
FT DISULFID 3461 3504
FT DISULFID 3490 3517
FT CARBOHYD 163 163
FT CARBOHYD 235 235
FT CARBOHYD 329 329
FT CARBOHYD 529 529
FT CARBOHYD 709 709
FT CARBOHYD 948 948
FT CARBOHYD 1409 1409
FT CARBOHYD 1479 1479
FT CARBOHYD 1523 1523
FT CARBOHYD 1530 1530
FT CARBOHYD 1625 1625
FT CARBOHYD 1751 1751
FT CARBOHYD 1988 1988
FT CARBOHYD 2088 2088
FT CARBOHYD 2089 2089
FT CARBOHYD 2507 2507
FT CARBOHYD 2642 2642
FT CARBOHYD 2679 2679
FT CARBOHYD 2748 2748
FT CARBOHYD 2762 2762
FT CARBOHYD 3069 3069
FT CARBOHYD 3194 3194
FT CARBOHYD 3232 3232
FT CARBOHYD 3545 3545
FT CARSPPLIC 485 1411
FT Missing (in isoform V1).
FT /FTId=VSP_003093.
DR 3562 AA; 388078 MW; 9BC565E8C1602D2 CRC64;
SEQUENCE

```


Db 2164 QSGEATLSTESFTKASVPTGKPEQVGRKTVMPASVHAYTABNELVTSEHDI 2223
QY 1941 PDGANTDEKLNPAI-ENTPVCFSNSLSLSDIDQNNKNEPIKETEP--DSQG 1997
Db 2224 TSLQTVTDTMEBEKAANELTFTVSFATNPLS-----EDVHSEWDRP-REILPKAIESG 2276
QY 1998 EPSK-PQASGYAPKSHVE-----DTPVCFGRNS----- 2026
Db 2277 EATEDFFISTQANHEHVELSVPTIRPHSEENKVEAESDEKILLFPNDRVTSAVIER 2336
QY 2027 --LSSIDS--BDDLQECISS-----AMPKKKPSRLKGDNEKHS-PRNMGIIGE 2074
Db 2337 KYLSSPTDTEEBELVQNIPTEDIPRLFTLPKEEPT-----NNELISDPLFSQGSQSD 2392
QY 2075 DLTL--DLKDIQRPDSEHGLSP-----DSENFDMKA-IOEGANSI 2111
Db 2393 EFTVPSVESLAVKETNTLSPWPHFHPASVGPVKLSTDKTVQVFESGTDNAGINEIITT 2452
QY 2112 VSLHQAAAAACLSRQASSDSILSKGISGLSPFHLTPDOEEKFTSNKGR----- 2166
Db 2453 AABELTETAYSMATSSPALBESSHNSKDKDITHYFLVIEDPYNKEMDHRGEGTSRP 2512
QY 2167 ILKPEKSTLETYK-----LESKIKGGKKVYKSLITCKVRSNSELGQ-- 2212
Db 2513 LPTFGDVSLESSHMLTTDDVTPVSVILSTPYLEMCKSLATSKMPRVRVLPESGEGS 2572
QY 2213 -----MKQLQANMPSISGRTHIHPGVNRSSSSTSPVSKKG 2250
Db 2573 GWDGVSDSPADTLTHTAPSVMEVELTASS-----HIPGVYSEVMIT----- 2615
QY 2251 PPLKTPASKSPSEG-QTATTPRGAKPSVKSLSVPARQTS-----QIGGSKKAPSRG-SR 2305
Db 2616 -----HVPGGSGQTVIT---GLASLFTTEKEEIVANRTAADPKGTSEELTSDTGMSL 2664
QY 2306 DSTPSRPAQPLRPIQSPGRNIGISPGNKLSPRTSSPSTASTKSSGSGKMS 2365
Db 2665 DIIPVDDRHLVINSVYGDILIEERLOI-PSEKTIIDMDHKSMPEDIIISVQTMEN 2723
QY 2366 YTPGGRMQOONTKQ-----TGLSKNVAISIPRSESASKG----- 2400
Db 2724 LVIRSTQVSDNNKAEDKYDILNFTSVENSFGSDNLSLTSIQPSSESVTAGHGPK 2783
QY 2401 LNQNNNGANKKVELSRMSTKSSGSESDRSPVLRQSTIKEAP---SPTLRKLE 2457
Db 2784 LVKDLGSGYAMOFATETLTT-----VLNELGIFLPTVPLSVPHMHEK 2830
QY 2458 ESASFES-----LSPSSRRPASPTRSQATPVLSPSLDMSLTHSV---QAGGWKRLPP 2509
Db 2831 ES-EFEAKHIGRTSTDDVVEPVTS-ANNQVITDQKTMISIGFSGMQGSESGDKKPMIP 2888
QY 2510 NLSP--TIEYNDGEPAPKHDIASHSPSPRLINSGTIWKREHSHKSSSLPRVSTWRET 2567
Db 2889 SLTFLDLMETEKALTDTDFDVSVMTTQSMSCATVSSSEEEKS-----TVYMQT 2939
QY 2568 GSSSILSASSESEKAKSDEKH-----VNSISGTQSKENQVSASKG 2610
Db 2940 KSASTEYEEDTSLNSVSNQKSSVTVLVNGVSKYVEVIIPTSSAKDSQSDHSSDG 2999
QY 2611 TWKIKENERS-----PNSTSTQTVSS--GATNGAESKTLIYQMAPAVSK-----TDDVVV 2659
Db 3000 TFEVSDMAATYKPPPTDITTVSLLVFPSPSESISTESTPHFNKFTVRSSETES 3059
QY 2660 RIEDCPT--NNRPSGRSP-----TGNTPPVIDSV-----SEKANPNIKDS 2697
Db 3060 SVNDLIIENATVSGDSFSDHYDPTAFWNFGERTSTDPKLSLIEVEFSERVKNSQES 3119
QY 2698 KDNQAKQVNGVSGVPMETVGLNRLNSF-IQVDAPOKQGTETKPGQNNVPVSETNESSI 2756
Db 3120 DRSTERERPLSLGAP-----VSDSPNSIEVGFKPQDEAVTMTLSTSLDLRSLETQSAL 3174
QY 2757 V-----ERTPFSSSSSKSHSPSGTVAARVTPFNYNRPSRKSSADSTS-----AR 2801

Db 3175 LGPLLQOQEITTISSNIATNTAPGN-----NPYSNEQSTISSSELLNTIELVTS 3223
QY 2802 PSQIPTPNNTKXKRDSDKTDSTESSGTQSP 2831
Db 3224 SFSUPEVTNGSDFLIGTSVGSVEGTAVQIP 3253
RESULT 12
PCLO RAT
ID PCLO RAT STANDARD; PRT; 5085 AA.
AC Q9JKS6; Q9JTLI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.",
RL Neuron 25:203-214 (2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4689; ASP-4674;
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP ALA-4694.
RX MEDLINE=21181819; PubMed=11285225;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.",
RL EMBO J. 20:1605-1619 (2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AF138789; AAF07822.2; -
CC EMBL; AF227534; AAF63196.1; -
CC HSSP; P04410; 1A25.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC GO; GO:0005509; F:calcium ion binding; IDA.

DR GO: GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR GO: GO:0005522; F:profilin binding; ISS.
DR GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO: GO:0016080; P:synaptic vesicle targeting; NAS.
DR InterPro: IPR000008; C2.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR008899; ZnF_piccolo.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF05715; ZF_piccolo; 2.
DR SMART: SM00239; C2; 2.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00499; C2 DOMAIN 1; 1.
DR PROSITE: PS00004; C2 DOMAIN 2; 2.
DR PROSITE: PS0106; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 372 491
FT ZN FING 523 547
FT ZN FING 1010 1033
FT DOMAIN 2351 2362
FT DOMAIN 4442 4536
FT DOMAIN 4653 4752
FT DOMAIN 4968 5059
FT VARSPLIC 4876 4880
FT VARSPLIC 4881 5085
FT MUTAGEN 4668 4668
FT MUTAGEN 4674 4674
FT MUTAGEN 4688 4688
FT MUTAGEN 4688 4689
FT MUTAGEN 4689 4689
FT MUTAGEN 4690 4691
FT MUTAGEN 4692 4693
FT MUTAGEN 4694 4694
FT SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
Query Match 2.7%; Score 390; DB 1; Length 5085;
Best Local Similarity 20.0%; Pred. No. 8.9e-07;
Matches 386; Conservative 241; Mismatches 775; Indels 532; Gaps 81;
QY 1187 IPSSOKQSFSSKSSGSSGSSSTSTPSSNAKRONLHPS--SAQSRSGOPQ 1244
DB 67 VPPAAESPMHRKQELDSSQAPQPGPPGPRTPQCLSKSRITDTFRSEQLGRSP 126
QY 1245 KAATKVSINQETITQYCVBDTPICFRCSLSLSAB-----DEIGNQNTQPADS 1298
DB 127 STISLKESKSRDTEKEEYKSNMPPGFFDVNPLSAVSVNKNPNFOLISDSASQ--ET 185
QY 1299 ANTLQIABIKIGTRAEADPV---SEVPASVQHPRTKSSRLQGS---SLSESARHKA 1352
DB 186 TKQKVVOKEQKSGMAKPLQQSPAPIPKQGVKEVIQDSSSPKSVSSQAEKVP 245
QY 1353 EFSGAKSPKSGAQTPTKSPPEHYVQETPLMFRSCTSVSSLDSPESRSIASSVQSEPCSG 1412
DB 246 QAPGTGKPSQSPAQTPE-----QASGPKFVAQP--G 277
QY 1413 MVSGIISFDLPDSQGTMPBSRSKTPPPQCTAQTKEVPKVA---PRAKRESGPKQ 1469
DB 278 SAKATVQPGPAKSPAQ--PAGTGKSPAQP-----AKTEGQAQGLEKTSSSQQPGPKS 329
QY 1470 AAVNAAVQORVQLP-----DADTLLHFAETSTPDGFCSSLSALSILDEPFIQKVLE 1523

DB 330 LAQTPGHGKPLGFKVSPAQPGTAKHPAQPGPQ-----TAAKVPGP----- 372
QY 1524 IMPVQENDNGNETESQPKESN-----ENQKEAEKXIDSEKLLDDSDDDDD 1572
DB 373 TKTPAQSGGFG-KTPAQPGPFTKPSQOPIPAKPPQCPQVATKTQPOQ----- 419
QY 1573 EILBECIISAMPTKSSKAKKPAQTASKLP-----PFVARKPSQLPVYKLLPSQNL 1624
DB 420 -----SAFAKPOQPAKAPQPPQPTPAKPPQPPPTPAKPPQ-----PTATKP 465
QY 1625 QPKHVSFTPGDDMPRYCVVEGTPINFSTATSLDITSPNELAAGEVGRGAQS--- 1681
DB 466 QPOPTATKPHHOQGLAKPSAQOPTKSIQTVTGRPLQPPPTS-AAQTPAQSLSTICP 524
QY 1682 --GEFERDRTIP-----TEGRSTDEAOGKTSSTVITPELDD-----NKASEGDI 1723
DB 525 LCNTTELLHIPEKANFNTCTEQSVCSLCGFNPNPHLTIKEWLCNLCMQQALGGDL 584
QY 1724 LAECINSAM--PKGSHKPRVKINDVQVQAASASSAPNKO-----LDGKK----- 1769
DB 585 AAAPSSPQPTPKAATAAPTATASKSPVPSQASPKKEPPSKQSPKALESKKPEPKPP 644
QY 1770 --KKPTSPVKP-----SSLDPDDDDVLSREKALRAKEN 1887
DB 645 EPKKPEPKPPPLVKQPTLHGPTATAPQLPVAEALPEPAPPEKPSGPIPEQAKAPVG 704
QY 1779 -----IPONTYRTRVRKNADSKNNLNAERVSNDKSKKNNKSKDFNDKLPNNEDRV 1834
DB 705 VEPKQPMETRADIQSSSTTKPDILSSQVQAQVKTASPLXTDSAKPSQSPPTGEKT 764
QY 1835 RGSFADPSHHYPIEGTPYCFERNDSL-----SSLDPDDDDVLSREKALRAKEN 1887
DB 765 -----TFLDSKAMPRAESDKIISQPGPSESNDPKHIDPIQKDEPFKA-QP 811
QY 1888 KESEAKVTSHTELTSNQOSANKTOAIKOPINRQPKLIQKOS-----TFPQS 1936
DB 812 KGSPKPEKVPKSGSPPTSGTRPTAGAAAPPSQPPK--QEQSRFSLNGLGITDAPKS 869
QY 1937 SKDIPURGAATDEKLFQNFALIENTVCFSHNSLSLSDDOENNNKENEPIKETE--PPD 1994
DB 870 QPTTPOE--TVTKLFGFGAS-----IFSQASNL--ISTAGQQGPHPTQPAAPSKQAPT 921
QY 1995 SQGEPSKPAQASGAPKSFHVEDTPVCF--RNSLSLSIDSDLLQECISSAMPKKK 2052
DB 922 SQ-----SPAAQGPAGKSTGQLPPAPAKATAVKKAKEAENLESKEPQAPTAKTEKDK 977
QY 2053 PSRLK-----GDNKXHSRNMGGIIGEDLTLDLK--DIQRPDSHGLSPDSSENFDKA 2103
DB 978 PPPAKVGPSPSPBPAPKPAH-----PDKTKPKPACPLCRTELNLG--SQEPNEN--T 1029
QY 2104 IQEGANSI-----VSSLHQAAAAACLSRQASS-----DSDSILSKSGISLSPF 2148
DB 1030 CTECKNQVCMGNFNTPHLTIQEWLNCQTRALSGQDGMGNKMPAPSP--PKASPM 1088
QY 2149 -----HLTP-----DOEKPFTSNKGRILKPGKEKSTLETTKIESES 2185
DB 1089 PAPAEPSSOKTPTGTQVKGKKEAEGTEAEKVPPEKETAIEKTPPMVTTDQKLESE- 1147
QY 2186 KGIKGGKVKYKSLITCKVRSNSE--ISGQVKQ-----PLQANPISRCRTMIHP 2234
DB 1148 -----GKKKGVSALEPKPSEEEKALSADKKEKPPAEKPPLEKPP----- 1190
QY 2235 GVRNSSSTSPVSKGPPPLKTPASKSPSEGTATTSPRGAKPSVKSELSVPAQTQSIGG 2294
DB 1191 -----IPVDKKLPPPEAKPLS--SEGE-----KHEILKAHVQIPEEP 1226
QY 2295 SVKAPSRGSRDTPRPAQOPLSRPIQSPGRNNSIPGRNGISPPNKLSQLPRTSPSTA 2354
DB 1227 TGVAAKAGEE-----QPDSP-----EALPGATPLTPKAGEKERAQAQAE 1272
QY 2355 STKSSGSKMSY-----TSPGQMSQNLTKOTGLSKNA-----SSIPRS 2394

Db 1273 GSKDQGGERSKEKTEKEDKSDTSSQQKPSQGLSDTGYSSDGISSGLGIPSLIPSD 1332
 QY 2395 E-SASKGLNOMNGNANKKVELSRMSYKSGESDRSERPVLVRQSTFIKEAPSPIL- 2452
 Db 1333 EKDLLKGL-----KKDSFQESS-PSGPSDLAKLESTVL-----SILEAQASTIV 1376
 QY 2453 RRLLESAPESIPSSRPASPTSCAQTPTVSPSLPDMSLSTHSSVOAGGWRKLPPLNS 2512
 Db 1377 GEAEKKTPOKLSLPEKQDQQTOTASETL-----DIIIS-EBEIKESQKK-----VS 1425
 QY 2513 PTTEYNDGRPAKHDIARGHSESP-----SRLPINRSGTWKREHSHKSSSLPRVSTWRT-- 2567
 Db 1426 PKKDSQGFPSR-----KEHKEKPELVDDLSPRASYSVDVEDSSESENSEPVVRRKRRTSI 1480
 QY 2568 GSSSS-----ILSASSESEKAKSEDEKHVNS-----ISGTKQSK 2602
 Db 1481 GSSSDSEYKQEDSQSGSEEBEDFIRKQIIEMSAD-EDASGSEDEEFIRKQIIEMSAD 1539
 QY 2603 ENOVSAKGTWRKI--KENEPSPTNISTQTVSSG-----ATNGA 2638
 Db 1540 KRE-EAKGKGVAGKHRLTRKSTSFDDAGRRHSHWDEDETPDESPELKFRETQSK 1598
 QY 2639 ESKTLIQMAPAVSKTEDVWRIEDCPINPRSGRPTGNTPPVIDSVSEKANPNIKDSK 2698
 Db 1599 ESEELVVAGGGLRRPKTIEL---NSTIADKYSESSQKKTILYFDDEPELEMESELTDSP 1655
 QY 2699 DNOAKQNVGNGVPMRTVGLNLSFIQVDPDQKGTBKPGQNNPVPVSETNNESSIVE 2758
 Db 1656 EDRSR---GGSSSLH-----ASSFTFG-TSFTSVSSLDEDS--- 1688
 QY 2759 RTPFSSSSSKHSPGTVAAVTPFNPNPSPKKSADTSAPRSQIPTPVNNNTKRRS 2818
 Db 1689 ---DSSPSHKKGESKQOKARHSHGFLPTIEDSSEEBELREBEL---LKEQEKQREL 1742
 QY 2819 KTDSTESSGTQSPK 2832
 Db 1743 EQQRKSSSKKSKK 1756

RESULT 13
 PCLO HUMAN
 ID PCLO HUMAN STANDARD; PRT: 5147 AA.
 AC Q9Y6V0; Q43373; Q60305; Q9BVCS; Q9UIV2; Q9Y6U9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Aczonin) (Fragments).
 GN PCLO OR ACZ OR KIAA0559.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE OF 1-759 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9439764; PubMed=10508862;
 RA Wang X., Kibechull M., Laue M.M., Lichter B., Petrasch-Parwez E.,
 RA Kilimann M.W.;
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rim and bassoon and binds
 RT profilin.";
 RL J. Cell Biol. 147:151-162(1999).
 RN [2]
 RN SEQUENCE OF 552-4404 FROM N.A.
 RA Kraemer J., Wollam C., Wohlmann P., McGrane B.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara C.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.

The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 DNA Res. 5:31-39(1998).
 [4]
 RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=2388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan B., Moore I., Max S.I., Wang J., Haieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [5]
 RP SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Elliott G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y19188; CAB60727.1; -;
 DR EMBL; AC004903; AAD20936.1; -;
 DR EMBL; AC004886; AAD21789.1; -;
 DR EMBL; AB011131; BAA25485.1; -;
 DR EMBL; BC001304; AAH01304.1; -;
 DR EMBL; AC004082; AAB97937.1; -;
 DR PIR; T00634; T00634.
 DR HSSP; P04410; 1A25.
 DR Genew; HGNC:13406; PCLO.
 DR MIM; 604918; -;
 DR GO; GO:0005856; C:cytoskeleton; NAS.
 DR GO; GO:0045202; C:synaptic junction; ISS.

DR GO: GO:0005509; F:calcium ion binding; ISS.
 DR GO: GO:0005544; F:calcium-dependent phospholipid binding; ISS.
 DR GO: GO:0005522; F:profilin binding; ISS.
 DR GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO: GO:0016080; P:synaptic vesicle targeting; ISS.
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR001565; Synaptotagmin.
 DR PRINTS: PR00360; C2DOMAIN.
 DR PRINTS: PR00399; SYNAPTOTAGN.
 DR SMART: SM00239; C2; 2.
 DR PROSITE: PS00499; C2 DOMAIN 1; 1.
 DR PROSITE: PS00044; C2 DOMAIN 2; 2.
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
 Repeat; Alternative splicing.
 FT NON_TER 1
 FT DOMAIN 400 465
 FT ZN_FING 499 523
 FT ZN_FING 959 992
 FT NON_CONS 1010 1011
 FT DOMAIN 2300 2325
 FT DOMAIN 4391 4442
 FT DOMAIN 4544 4633
 FT DOMAIN 5031 5121
 FT VARSPLIC 4404 4404
 FT FT
 FT VARSPLIC 4534 4534
 FT VARSPLIC 4576 4576
 FT VARSPLIC 4757 4761
 FT VARSPLIC 4762 5147
 FT SEQUENCE 5147 AA; 563537 MW; CD5D84390498CD3C CRC64;
 Query Match 2.7%; Score 389.5; DB 1; Length 5147;
 Best Local Similarity 19.4%; Pred. No. 9.4e-07;
 Matches 508; Conservative 313; Mismatches 959; Indels 837; Gaps 122;
 718 IAMGSAALRLMANRPAKYKIDANIMSPGSSI-----PSLHVRKQKALEELD 765
 1 IPAGNEADLSQI--SEERRQIAVMSRAQGLPKGSVPPAAEAFSMH-RKQ-----ELD 52
 766 AQHLSSETFDNIINSPKASHRSKQHKOSLYGVYFDNRHDDNRSDNFTGNMTVLSPY 825
 53 SSH-----PPKQSGRPP-----DPCRPAQPLGSKSRITDTF 83
 826 LNTTVLPSSSRGSLDSRSSEKORSLEERIGLGNVYHPATENPG-TSSKRGLQISITA 884
 84 RSEQLPGRSFTSLKESKRTD-----LKEEHKSMMPGFLSEYNALSAVSV 133
 885 AQIAKWEVSAIHVSQEDRSSGSTTELHCVTDE----- 918
 134 VNKENPFLISDSEASQEE-----TTKQKVQKEQKGEGIIKEPQQQPPKPIPKQQ 188
 919 --RNALRSSAAHTSNYNTKSENSNRCTCMP-----YAKLEYKRS 959
 189 PCRDLQDGTPKSIS-----SQQPEKIKSQPPGTGKPIQGTPTQTDHAKLPLQRD 241
 960 SNDSLNSVSSDGYKRGQ-MKPSIESYSEDESKEFCSYGOVPADLAHKIHSANHMDND 1018
 242 A--SRPQTQADIV--RSESVKPSLPSKSPPI-----OQPTGKPPAQOQHEKSKP 290
 1019 GELDTPIPNYS-----LKYSDQLNSGRQSPQNERWAPKHIIEDEIKSQSQRNRQST 1073
 291 GPAPKPPAQSGLTKPLAQOQGTGVKPPVQPPGTGTPKPPAQP---LGPAPKPPAQQTSEKPS 347
 1074 TYPVVTSTDDHLKFKQHFQOQECVSPYRSGANGSETNVRGNSGNCINQVNSLQED 1133
 348 EQP-----GPKALAQPPGVSK-----TPAQQPGPAKPTQOVGT-----PKPLAQQP 389

QY 1134 DYEDDK-----PTNYSEYSEBQHEBERPTNYSIKNEEKRVHDQPIDYSLKYATDIPS 1189
 Db 390 GLOSPAKAPGTTTPTAQTKPSPQPGSTKPPQ--QFGPAKPSPPQ-----GSTKPPS 441
 QY 1190 SQKQSFSSKSSGQSKTEHSSSENTSTPSNAKRONQLHPSSAQSSGQPKAATC 1249
 Db 442 QQPGS---AKPSAQPSPAK---PSAQPTKPVSTQFGKPLQPPVPSPAKAPP----- 490
 QY 1250 KVSINQETIQTVCYVEDTPIC-----FSRCSLSLSLSAEDEICGNQTOE 1295
 Db 491 -----SQLEKTKIC-----PLCNTTELLHVPEKANFNTECQTCTVCSLCGFNPHELTE 541
 QY 1296 ADSANTLQIAIEIKELIGTRSAEDPVSEVPAVSHBRTKSSRLOGSSLSSESARHKAVEPS 1355
 Db 542 AKWLCNLN-CQMKALGGDLAPVSPPOPKLKTAPVTTTSVAVSKSPQOQTSPK---K 596
 QY 1356 SGAKSPSKGAQTPKGPPEHYVQETPLMFSRCTSVSLDSFESRSTASSVQSPSCGMWS 1415
 Db 597 DAAPKQDLSKAPEKPKPPLVKQPT----- 623
 QY 1416 GILSPDLPSQGTMPPSHESKTPPPPPQTAQTKREVKNKAPTAKRESGPKQAAVNA 1475
 Db 624 G--SPSAKAKOP-----PEADLSKPAPPEPSVPSE--QDKAPVADKKPKQPKM----- 669
 QY 1476 VQRVQVLPDADTLLHFATESTPDGFCSSSLSALSLEDEPIQKDVLRIMPVQVENDNGN 1535
 Db 670 -----VKPTTLVSSSATTXPD:IPSSKVQSOABEKTTPPL----- 705
 QY 1536 ETSEQPKES-----NENOEKAEKTIIDSEKOLLDD 1566
 Db 706 KTDSAKPSQSPPTTGEKVTFFDSKAIPRASDSKIIISHPGPSSSESGKQKV----- 756
 QY 1567 SDDDDIEILIECIISAMPTKSSKAK-----KPAQTASKLPP---PVARK---PSQLPV 1614
 Db 757 ---DPVQKKEE-----PKAQTKMSPKPDAPKMPKSGTTPPPRPTAGTGTTPQOSPK 807
 QY 1615 -----YKL-----LPSQNELQPKHVS-----FTPGDDMPRVY 1642
 Db 808 PQBSRRFSLNLGSIITDAPKSTPTTQETVTGKLFQFGASIFSOASNLISTAQCPGHSQ 867
 QY 1643 CVRGTPINFTATSLDOLITESPENELAGEVGRGAQSGEFKRTDITPEGRSTDEAQ 1702
 Db 868 SGFGAPWKQAPAS-QPPTSQGPBK--STGQAPPAKSIPIVKETKAPAAEKLEPRAEQ 924
 QY 1703 GKTSSVT-----IPELDDNKAEGDILAEICINSAMPK--GKSHKPPRVKIMDQVQASA 1755
 Db 925 APTVKRTETETKPPKIDSKS-----LTAPQKAVLTKLEKSPKPESTCLC-KTELNI 979
 QY 1756 SSSAPNKN-----QLDGKKKKPTSP-----VKP 1778
 Db 980 SKDPPNFNTCTECKQVNCNLCGFNPTPHLTENCQTORA:SGQLGDIRKMPAPSGPKASP 1039
 QY 1779 IPQNTYVTR-----VRKNADS-----KNLNAER-----VFSNKNDS 1811
 Db 1040 MPVPTSSSQKTAVPQVKLVKQEQEVTEAEKVLKVKETLSMEKIPMWTTTQKOE 1099
 QY 1812 KQNLKNNKDFNDKLP-----NNEDVRGS-----FAFDSPHHYTPIEG--TPYCF 1856
 Db 1100 ESKLEKDKASALQEKKPLPEEKLIPEEEKIRSEKKPLLEKPP---TPEDKLLPEAK 1156
 QY 1857 SRNDSLSSLDFFDDVDLSREKAEALKA---KENKESAK---VTSHTELTSNQQSANK 1309
 Db 1157 TSAPEQKHDLKLSQVIAEEKLEGVAVKTVQEGKQPTQKMEGLPSGTPQSLFKEDDKT 1216
 QY 1910 TQIAKQPINRGPKFILOKQ-----STFPQSSKDIPIRGAA----- 1946
 Db 1217 TKTIKEQPQPCPTAKPDQEKEDDKSDTSSSQPKSPQGLSDTGYSSDGISSSLGIETPSLI 1276
 QY 1947 TDE-----KLONFAIENTVCFSHNSLSL-----LSDIDQENNKENEP--- 1986
 Db 1277 PTDEKIDILKGLKDKDSFSQSSSPSPDLAKLSTVLISLEAQASTLADESEKKTQFHEV 1336
 QY 1987 -----IKETFPDPSQGEPSK-----PQASGVAPKSFHVE 2015

```

Db      1337 SPEQPKQDKQKTSLETLEITISEEBIKESQBERKDTFKDKSQDDIPSSKDKHEKSEFVD 1396
Qy      2016 DTFVCFSRNLSLSLSDGDLQICISAMPKKKPSRL-KGDNKHSPPNMGJILGE 2074
Db      1397 DIT---TRRPYDSVESSE-----NSPVQKARTSVGSSSDSEYKQSDSQSGSEE 1447
Qy      2075 DLTLDKDIQDPSEHGLSPDSEFNPKAIQE-----GANSIVSSLIHQAAA 2120
Db      1448 EDFIRKQIEMSADEADASGEDEDFIRNOLKEITSSSTESQKBEETKGGKITAGKHR-- 1505
Qy      2121 AACLRCASDSDSILSLKSGISLGFPHLTPDQEKPTSNKPRILKPGKSTLETWK 2180
Db      1506 ---LTKRSSTIDEDAGRR-----HSHWDEDEAF--DESP-----ELKRETKS 1545
Qy      2181 IESKSGIK--GGKVKYKSLITGVRSNBEI-----SQMKQPL-----QANPMSIS 2225
Db      1546 QSEELVVTGGGLRRPKT-----IELNSTIADKYSAESQKTSLYFDEEPELEMESLT 1600
Qy      2226 -----RGRWTH-----IPGVNRS-----SSTSPVKKGPPKTPASKSPSEG- 2264
Db      1601 DSPEDRSRGEGSSSLHASSFTPGTSTPSVSSLEDSDSPSHKKGSKQQRKARHPGP 1660
Qy      2265 --OTATTPRGAKPSVKSLSPVAROTSGTSGSKAPSRSGSDSTPSRPAQOPLSRPTQ 2322
Db      1661 LLPFIEDSSEBELREBEELLKQEKQRETEQQRKSSSKKKDKDELRAQRRRERPKT 1720
Qy      2323 SPGRNSTSPGRNGISPPNKLQ-----LPRTS-SPSTASTKSGSGKMSYTSIPGRMSQ 2375
Db      1721 PP--SNLSPIEDA-SPTTEURLQAAEMELHRSCSYSPSIESDPEG--FEISPEKIEV 1775
Qy      2376 QNLTK-QTGLS-----KNASSIPRSEGASKGL-----NOMNGN-----GANKKVEL 2416
Db      1776 QKVTKLPTAVLSPTDEQIMKESQKALKSAEMYEMMKTKYKAFPAANERDEV 1835
Qy      2417 -----SRMSSTKS-----SGSSSDRSE-----RPVLVROSTFTK 2445
Db      1836 FEKEPLYGGMLEIDYIESLVEDTYNGSVGSLTLTQEEENGFMQKGREQKIRLSEQIV 1895
Qy      2446 EAPS-----PTLRKLEESAFESLSPSRPASPTSOAQPTVLSPS--LDMSLSTHSSVO 2500
Db      1896 EDPMQKITDQKFEYELSHSVVP-----QEDIVSSSTIIFE-----SHEIVD 1939
Qy      2501 AGGW-----RKLPNLSPTIEYNDGRPAKHDIARSHSESPSRLPINRSGTWRKREHSK 2553
Db      1940 LGTMVSTEEERKL---LDADAAYEE--LMKROQMQLTPGSSPTQAFIGEDMT---ESTM 1991
Qy      2554 HSSSLPRVTRWRTGSSSSILSASS-----ESSEKAKSEDEKHNSISGTQSKENQVYSAKG 2610
Db      1992 DFDMPDASL-----TSSVLGASLTDSTSSATLSIPDVKITQHFSTBEIEDEYVT--D 2043
Qy      2611 TWRIKE-----NEF-----SPTNSTSQTVSSGATNGAESK-TLIYQMAP 2649
Db      2044 YTEIQIIAHESILITYSPSSSATSVPSPDPSLTSSVSVCTTSSDSSPITLDSIT 2103
Qy      2650 AVSKTEVDVRIEDCP-----INPNRSGRSPGTNTPP----- 2681
Db      2104 VYTEPVDMTKFEDESEIBSSSTYPGSGIIDYPEISVSLDRTAPPDGRASADHIVISLD 2163
Qy      2682 -----VIDSVSEKANPNTKDS-----KD--NOAKQNVGNSV-----PMTVGL 2718
Db      2164 MASSIIESVVPKPGPVADTVSTDLTISEKDPVKKAKETGNGIILEVLEYEDKKELEA 2223
Qy      2719 ENRLNSTIQV--DAPDQKGTEIKPGQNNVPVSETNESSIVERTPFSSSSSS-----SKHSS- 2772
Db      2224 ERTKSSLSSETVFDHP-----PSSVIALPMKEQ-----LSITYFTSGTFTGQKQPASQ 2270
Qy      2773 -PSCGTVAAVTPFNYNPSPKSSADSTSAARPSQITPT 2808
Db      2271 LPSSGSPSVSLPAKPRBFRFRSSSLD-ISAQPPPPPPP 2306

```

RESULT 14

```

MLL3 HUMAN
ID MLL3 HUMAN STANDARD; PRT; 4911 AA.
AC Q8NEZ4; Q8NC02; Q8NDF6; Q9H9P4; Q9NR13; Q9P222; Q9UDR7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-
DE lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43)
DE (Homologous to ALR protein).
GN MLL3 OR HALR OR KIAA1506.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal thymus;
RX MEDLINE=21888622; PubMed=11891048;
RA Ruault M., Brun M.-E., Ventura M., Roizes G., De Sario A.;
RT "MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a
RT chromosome region frequently deleted in myeloid leukaemia.";
RL Gene 284:73-81(2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21574953; PubMed=11718452;
RA Tan Y.C., Chow V.T.;
RT "Novel human HALR (MLL3) gene encodes a protein homologous to ALR and
RT to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated
RT with leukemia and developmental defects.";
RL Cancer Detect. Prev. 25:454-469(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22737999; PubMed=12853948;
RA Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
RA Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
RA Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
RA Fewell G.A., Delehaunty K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
RA Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
RA Vanbrunt A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
RA Ozersky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
RA Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
RA Kozlowski-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
RA Tin-Wollam A.-M., Abbott A., Minx P., Maupin R., Strommatt C.,
RA Latreille P., Miller N., Johnson D., Murray J., Woessner J.F.,
RA Wendt M.C., Yang S.-P., Schultz B.R., Wallis J.W., Spieth J.,
RA Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,
RA Hickenbotham M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
RA Clifton S.W., Chissoe S.L., Marra M.A., Raymond C., Haugen E.,
RA Gillett W., Zhou Y., James R., Phelps K., Iadonato S., Rubb K.,
RA Simms E., Levy R., Clendenning J., Kaul R., Kent W.J., Furey T.S.,
RA Baertsch R.A., Brent M.R., Keibler E., Flück P., Bork P., Suyama M.,
RA Bailey J.A., Portnoy M.E., Torrents D., Chinwalla A.T., Gish W.R.,
RA Eddy S.R., McPherson J.D., Olson M.V., Eichler E.B., Green E.D.,
RA Waterston R.H., Wilson R.K.;
RT "The DNA sequence of human chromosome 7.";
RL Nature 424:157-164(2003).
RN [4]
RP SEQUENCE OF 556-3865 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [5]
RP SEQUENCE OF 3193-3865 AND 4460-4911 FROM N.A.
RC TISSUE=Placenta;
RA Isozai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

```

Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Rana K.K., Iwayanagi T.;
 "NEDO human cDNA sequencing project.";
 Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SEQUENCE OF 3879-4911 FROM N.A.
 RC TISSUE=Testis;
 RA Duesterhoeft A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 [7]
 RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX (ISOFORM 2).
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=22371496; PubMed=12482968;
 RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
 RA Kwak E., Barlow N.A., Berger S.L., Chow V.T., Roeder R.G.,
 RA Acorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
 RA Lee J.W.;
 RP "Activating signal cointegrator 2 belongs to a novel steady-state
 RC complex that contains a subset of trithorax group proteins.";
 RL Mol. Cell. Biol. 23:140-149(2003).
 CC -1- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a
 CC coactivator complex of nuclear receptors, involved in
 CC transcriptional coactivation. MLL3 may be a catalytic subunit of
 CC this complex, which weakly methylates Lys-4 of histone H3. This is
 CC a specific tag for epigenetic transcriptional activation. May be
 CC involved in leukemogenesis and developmental disorder.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
 CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
 CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
 CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
 CC MLL2 and MLL3, and ASH2/ASCL2. Interacts with histone H3.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS;
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q8NEZ4-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8NEZ4-2; Sequence=VSP 008561, VSP 008562;
 CC -1- TISSUE SPECIFICITY: Highly expressed in testis and ovary, followed
 CC by brain and liver. Also expressed in placenta, peripheral
 CC blood, fetal thymus, heart, lung and kidney. Within brain,
 CC expression was highest in hippocampus, caudate nucleus, and
 CC substantia nigra. Not detected in skeletal muscle and fetal liver.
 CC -1- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B
 CC and H4, and may have a H3 lysine specific methylation activity.
 CC -1- MISCELLANEOUS: Found in a critical region of chromosome 7, which
 CC is commonly deleted in malignant myeloid disorders. Partial
 CC duplication of the MLL3 gene are found in the juxtacentromeric
 CC region of chromosomes 1, 2, 13 and 21. Juxtacentromeric
 CC reshuffling of the MLL3 gene has generated the BAGE genes.
 CC -1- SIMILARITY: Belongs to the TRX/MLL family.
 CC -1- SIMILARITY: Contains 1 DHHC-type zinc finger.
 CC -1- SIMILARITY: Contains 6 PHD-type zinc fingers.
 CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -1- SIMILARITY: Contains 1 SET domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaborat
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstat
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no
 CC modified and this statement is not removed. Usage by and for commerc
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announ>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AY024361; AAK00583.1; -
 DR EMBL; AF264750; AAF74766.2; -
 DR EMBL; AC006017; AAD45822.1; -
 DR EMBL; AC104692; -; NOT ANNOTATED_CDS.
 DR EMBL; AC005631; -; NOT ANNOTATED_CDS.
 DR EMBL; AB040939; BAA96030.2; -

DR	EMBL; AK022687; BAB14179.1; -
DR	EMBL; AK075113; BAC11409.1; -
DR	EMBL; ALG33324; CAD38780.1; -
DR	Genew; HGNC:13726; MLL3.
DR	MIM; 606833; -
DR	InterPro; IPR000637; AT hook.
DR	InterPro; IPR003889; FYrich C.
DR	InterPro; IPR003888; FYrich N.
DR	InterPro; IPR000910; HMG 12_box.
DR	InterPro; IPR003616; PostSET.
DR	InterPro; IPR001214; SET.
DR	InterPro; IPR001534; Znf_DHHC.
DR	InterPro; IPR001965; Znf_PHD.
DR	InterPro; IPR001841; Znf ring.
DR	Pfam; PF00505; HMG box; 1.
DR	Pfam; PF00826; PHD; 6.
DR	Pfam; PF00856; SET; 1.
DR	SMART; SM00542; FYRC; 1.
DR	SMART; SM00541; FYRN; 1.
DR	SMART; SM00398; HMG; 1.
DR	SMART; SM00249; PHD; 8.
DR	SMART; SM00508; Postset; 1.
DR	SMART; SM00317; SET; 1.
DR	PROSITE; PS00354; HMGI_Y; 1.
DR	PROSITE; PS00868; POST_SET; 1.
DR	PROSITE; PS00820; SET; 1.
DR	PROSITE; PS00216; ZF_DHHC; 1.
DR	PROSITE; PS01359; ZF_PHD 1; 5.
DR	PROSITE; PS0016; ZF_PHD 2; 6.
DR	PROSITE; PS00089; ZF_RING 2; 1.
KW	Transferase; Methyltransferase;
KW	DNA-binding; Nuclear protein;
KW	DNA-binding; Transcription regulation;
KW	Coiled coil;
KW	Zinc-finger; Repeat; Polymorphism.
FT	Zn_FING 341..391
FT	PHD-TYPE 1.
FT	RING-TYPE.
FT	ZN_FING 344..389
FT	RING-TYPE.
FT	PHD-TYPE 2.
FT	DHHC-TYPE.
FT	ZN_FING 436..489
FT	PHD-TYPE 3.
FT	ZN_FING 464..520
FT	PHD-TYPE 4.
FT	ZN_FING 957..1010
FT	PHD-TYPE 5.
FT	ZN_FING 1007..1057
FT	PHD-TYPE 6.
FT	DOMAIN 1084..1139
FT	POST-SET.
FT	DOMAIN 4770..4891
FT	DOMAIN 4895..4911
FT	COILED COIL (POTENTIAL).
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 644..672
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 1338..1366
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 1724..1787
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 3034..3081
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 3173..3272
FT	COILED COIL (POTENTIAL).
FT	DOMAIN 3391..3433
FT	A.T HOOK (BY SIMILARITY).
FT	GLN-RICH.
FT	DOMAIN 1719..1796
FT	PRO-RICH.
FT	DOMAIN 1834..2281
FT	PRO-RICH.
FT	DOMAIN 2412..2630
FT	ASP-RICH.
FT	DOMAIN 2690..2786
Query Match	2.6%; Score 380.5; DB 1; Length 4911;
Best Local Similarity	17.5%; Pred. No. 1.9e-06;
Matches 624; Conservative 505; Mismatches 1291; Indels 1153; Gaps	
QY	124 VNGSRESGT-----YLSELEKER-----SLILA-----DLKE-EKEDV
Db	566 VNGQSTPGIVPDVAQVHTTEQQKHPSLSLDTSLIAVSSQHTVTWELEKXISNEVD
QY	158 YYAQQLNTKRI-----DSLPLTFNFSLQTDTRRQLEYEARQIRVAMEEQLGCTQD
Db	626 EDLKMSSEVKHI CGEDQIEDKWEVTN-----IEVVTHQITV-QQEQLQLLEE
QY	211 EKGAQRRIARTQIQEKILRI-----ROLLOSQTAEARSQNKH
Db	674 ETVWSRESRPKLMBSVTLPLETIVSPHEESISLCPEEQIIVRLQGEKEQENSEL
QY	252 TGSHDAERQNGQGVGINMATSGNGGSTTRMDHETASVLSS-----SSTHSAP

CC -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus);
 CC reorganization to the kinetochore/centromere (coronal surface of
 CC the outer plate) and the spindle during mitosis.
 CC -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
 CC -!- PTM: Hyperphosphorylated during mitosis.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U19769; AAA82889.1; --
 CC EMBL; U30872; AAA82935.1; --
 CC EMBL; U25725; AAA86889.1; --
 CC PIR; PC4035; PC4035.
 CC Genew; HGNC:1857; CENPF.
 CC GK; P49454; --
 CC MIM; 600236; --
 CC GO; GO:0005699; C:kinetochore; TAS.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC GO; GO:0005819; C:spindle; TAS.
 CC GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
 CC GO; GO:0007088; P:regulation of mitosis; TAS.
 CC Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
 CC Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
 CC Lipoprotein; Prenylation.
 CC FT DOMAIN 14 197 COILED COIL (POTENTIAL).
 CC FT DOMAIN 273 769 COILED COIL (POTENTIAL).
 CC FT DOMAIN 823 1328 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1642 1746 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1862 2987 COILED COIL (POTENTIAL).
 CC FT DOMAIN 2207 2568 2 X 177 AA TANDEM REPEATS.
 CC FT REPEAT 2207 2386 1.
 CC FT REPEAT 2389 2568 2.
 CC FT DOMAIN 3015 3032 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT LIPID 3207 3207 S-farnesyl cysteine.
 CC FT VARIANT 3202 3202 K -> N (in dbSNP:7289).
 CC FT /FTId=VAR_014839.
 CC FT T -> A (IN REF. 2).
 CC FT L -> Q (IN REF. 2).
 CC FT G -> D (IN REF. 2).
 CC FT MISSING (IN REF. 2).
 CC FT MISSING (IN REF. 2).
 CC FT MISSING (IN REF. 2).
 CC FT V -> A (IN REF. 2).
 CC FT V -> L (IN REF. 2).
 CC FT ER -> DG (IN REF. 3).
 CC FT L -> Q (IN REF. 3).
 CC FT D -> N (IN REF. 2).
 CC FT ELNRAVAALNDQEAQ -> SSMRWQPCIMTKPEVS
 CC (IN REF. 3).
 CC SEQUENCE 3210 AA; 367589 MW; 11D8332496024334 CRC64;

Query March 2.68; Score 379.5; DB 1; Length 3210;
 Best Local Similarity 17.94; Pred. No. 1.3e-06;
 Matches 574; Conservative 502; Mismatches 1243; Indels 895; Gaps 138;
 7 DQLLKQVEALKWENSLRQLEEDNSHLTKLETEASNMKEVLKQLOGSIEDEAMSSGQI 66
 475 NELRRSEEMKKNLLKSHSEKAREVCHLEAEKLNKQCLNQSNFAEMKAKNTSQE 534
 67 DLLRLKE-LMLDSNFPGLKRSKMSLSRYSGREGSVSRSGECSPVPMGSPFRGFVN 125
 535 TMLRDQEKINQENSLTLEKL-KLAVADL-----EKQDCSQDLL----- 574
 126 GSRESTGVLEL-----EKERSLLADLDKBEKMDWYAYQQLNLTKEIDSLPLTENF 178
 575 --KKREHHIQNDKLSKTESKALLSALELKKE---YEELKE-EKTLFSCWKSENE 627
 179 SLQDTMTRQLVEARQIRVAMESQLQTC-QDMKRAQRRIARLQIQIEK-----ILR 230

Db 628 KLLTQM---ESEKENLQSKI---NHLETCLTKQOIKSHNEYNERVTRLEMDRENLSVEIRN 681
 QY 231 IROLLQSQATEARSS-----QNKHETGSHDAERQNEGQVGEINWATSGNQGGTTRM 284
 Db 682 LHNVLDSKSEVETQKLYANMELQOKAEFSQKQKQKIE-----NNCLK----- 724
 QY 285 DHETASVLSSSSTHAPRRLTSHLGTKVMYVLLSMLGTH--DKDDMGTLLAMSSQD 342
 Db 725 -----TSQLTGVDELEHKLQLLSNEIMDKRCYCQDLHAYES-- 762
 QY 343 SCISMROSGLPLLIQLLHNDKDSVLLGNSRSKARA-RASAALH-----NIHSQPD 397
 Db 763 -----LRDLL--KSDASLVTTNEDHQSLLAFDQPPAMHSHFANIIIGQGM 807
 QY 398 KRGRREIRVLHLLQIRAYCETCWEAHEPGMDQKNFM-PAPVEHQICPAVCVLMKL 456
 Db 808 PSESECR-----LEADQSPKNSAILQNRV-----DSL 835
 QY 457 SFDEEHRHAMNELGGLQIAELLOVDCMTGLTNDHYSITLRRYAGWALTNLTFGDVANK 516
 Db 836 EFSLESQKQMN-----SDLQKQCE-----ELVQIKGEIEENLMKABQMHO 875
 QY 517 ATLCMKGCVRAVLAQLKSEEDLQOVIASVLRNLNWRADVNSKKTLLREVGSKVLMMECA 576
 Db 876 SFVAETS-----QRISKQEDTSAHONVVAETLSALE-----NKEKELQLLNDKVEITEQAE 926
 QY 577 LEVKKEST--LKSVLSALWNLSAHCTENKADI CAVDGALAFVLGTLTYRSQTTLAIES 634
 Db 927 IQELKXSNHLLDSLELQLLS-----ETLSLEK- 955
 QY 635 CGGILRVSSLIATNE-DHRQILRENNCLQTLLOLHKLKSHLSITIVSNACGLTWLNSARNPK 693
 Db 956 -----KEMSSIISLNKREIEELTOENGTLKEINASLNQEKWNLOKS--ESFANYIDEREK 1009
 QY 694 DQEALWDMGAVSMKLNLIHSHKMIAMGSAALRNLMANRPKAKYKDIANIMSPGSSPSLH 753
 Db 1010 SISELSDQ-----YKQELILLQRCBETGNAYEDLSQYKAAQ----- 1047
 QY 754 VRKQKALAEALDAQLSETFDNINLSPKASHRSKQSHKQSLGYDYVFDNRHDDNRSN 813
 Db 1048 -EKNSKLECLLN--ECTSLCENKRNLEQLKEAFKQHEQFL----- 1086
 QY 814 FNTGNTVLSPLYNTTVLPSSSSRSGLSRSSEKDRSLERERIGLGNHYHPATENPGTS 873
 Db 1087 -----TKLAFAEERQNLMLLELTVOQALRSEMTDNQNNKSEAG----- 1126
 QY 874 SKRGL--QISTAAQIAKMEVEVAIHTSQDRSSGTTTELHCVTDERNALRRSSAAHTH 931
 Db 1127 --GLKQIMTLKBEQNKQKQVNDLLOENBQLMKVMKTKHECQNLSEPIRNS----- 1177
 QY 932 SNTYFTKSENRTCSM-PYAKLEYKSSNDLSNS-VSSSDGYGKRGKQMPKPSIESYSED 989
 Db 1178 -----VKRESERNQCCFKPQMDLEVKELISLDSYNAQLVQLEAMLRNKELK-----LQES 1227
 QY 990 DESKFCSYGOYPADLAHKHS-----ANHMD-----NDGELDTPIINSLKYSD 1034
 Db 1228 EKEKEC-----LQHELQITIRGDLTSLNLDQMOSQOBIISGLKDCED-----AEE 1270
 QY 1035 QLNSG--RQSPQNERWARPKHI-----IEDEIKQSEQRQSRNQSTTYPVYVTESTDDKH 1086
 Db 1271 KYISGFHELSTQNDN-----AHLQCSLQTTWNKLNLEKICEILQAEKYLVLTLNDR- 1325
 QY 1087 LKFPQHQCCQCVSPYRSGANGSETNRVSGNHGNCNVQSLCOEDDYEDDKTNTYSE 1146
 Db 1326 -----SECITATRK-----MABEVGKLINLVKILNDDSGLLHGE 1360
 QY 1147 YSEBQHEERPTNYSIKYNEEKRVHVDQPTDYSIKYATDIPSSQKQSFSPK--SSSQS 1205
 Db 1361 VEDIPGGFSGQF-----NEQHPVSLAPLDEGNSYEHLLTSLDKVQMFAELQKFLS 1413
 QY 1206 SKTEH-----MSSSENTSTPSSNAKRNQHPSSAQSRSGQ----- 1242
 Db 1414 LOSEKILHDQHCQWSSKWSLEQTVVDSLKAENLVLTNLNRFNFOCDLVKEMQLGLEGLV 1473

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:05:26 ; Search time 183.5 Seconds
(without alignments)
4888.383 Million cell updates/sec

Title: US-09-442-489f-7
Perfect score: 14566
Sequence: 1 MAASVQLLKQVLAALXNEN.....ESSCTQSPKHSGLYVTSV 2843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_todent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10544	72.4	2829	13 P70039	P70039 xenopus lae
2	5162	35.4	1056	11 Q8BNP7	Q8bnp7 mus musculus
3	3512.5	24.1	2274	11 Q9ZIK7	Q9zlk7 mus musculus
4	3415.5	23.4	2303	4 Q95996	Q95996 homo sapien
5	2232	15.3	489	11 Q8BD8	Q8brd8 mus musculus
6	2170.5	14.9	733	4 Q9UBZ1	Q9ubzi homo sapien
7	1991.5	13.7	1246	4 Q9Y632	Q9y632 homo sapien
8	1780.5	12.2	2416	5 P91667	P91667 drosophila
9	1765	12.1	2417	5 Q9VAS9	Q9vas9 drosophila
10	1705	11.7	1685	4 Q9UEM8	Q9uem8 homo sapien
11	1252	8.6	324	11 Q8C493	Q8c493 mus musculus
12	1223.5	8.4	1067	5 Q9Y1T2	Q9ylt2 drosophila
13	1221.5	8.4	1067	5 Q961B0	Q961b0 drosophila
14	1041	7.1	208	4 Q9P119	Q9p119 homo sapien
15	756	5.2	159	4 Q7Z2Q8	Q7z2q8 homo sapien
16	680	4.7	146	11 Q8C919	Q8c919 mus musculus

17	660.5	4.5	16223	5	Q8IR22	Q8ir22 drosophila
18	649	4.5	5327	5	Q76891	Q76891 drosophila
19	647.5	4.4	5412	5	Q9W596	Q9w596 drosophila
20	564.5	3.9	2310	16	Q8CMU7	Q8cmu7 staphylococ
21	554	3.8	2296	4	Q9UHA8	Q9uha8 homo sapien
22	543	3.7	2607	11	Q8BTI8	Q8bti8 mus musculu
23	541	3.7	108	4	Q9HAM6	Q9haw6 homo sapien
24	537	3.7	2768	5	Q9VC00	Q9vc00 drosophila
25	528.5	3.6	2283	2	Q8VQ99	Q8vq99 staphylococ
26	527.5	3.6	3257	5	Q9V736	Q9v736 drosophila
27	526	3.6	2752	4	Q9UQ35	Q9ug35 homo sapien
28	517	3.5	2271	16	Q99QY4	Q99qy4 staphylococ
29	515.5	3.5	2275	16	Q8NUJ3	Q8nuj3 staphylococ
30	513	3.5	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
31	511.5	3.5	1186	5	Q21227	Q21227 caenorhabdi
32	510	3.5	3392	3	Q7ZA38	Q7za38 ashbya goss
33	510	3.5	4322	10	Q7XXN1	Q7xxn1 oryza sativ
34	509.5	3.5	1188	5	O62302	O62302 caenorhabdi
35	509	3.5	3381	2	Q9KX33	Q9kx33 streptococ
36	505	3.5	18519	5	Q8ISF6	Q8isf6 caenorhabdi
37	505	3.5	18534	5	Q8ISF7	Q8isf7 caenorhabdi
38	498.5	3.4	7210	5	Q9V7G8	Q9v7g8 drosophila
39	498.5	3.4	9270	5	Q8MLD9	Q8mid9 drosophila
40	492	3.4	3111	5	Q9VH10	Q9vhl0 drosophila
41	489	3.4	3072	2	Q939N5	Q939n5 streptococ
42	486	3.3	2178	2	Q9KWR3	Q9kwr3 streptococ
43	483.5	3.3	3443	11	Q8JZM8	Q8jzm8 mus musculu
44	483	3.3	2910	10	Q9FNDS	Q9fnd5 arabidopsis
45	477.5	3.3	3241	5	Q7YZH1	Q7yzh1 drosophila

ALIGNMENTS

RESULT 1

ID	P70039	PRELIMINARY;	PRT;	2829	AA.
AC	P70039;	P79934;			
DT	01-FEB-1997	(T-EMBLrel. 02, Created)			
DT	01-MAY-1997	(T-EMBLrel. 03, Last sequence update)			
DT	01-OCT-2003	(T-EMBLrel. 25, Last annotation update)			
DE	Adenomatous polyposis coli.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Vlemrickx K., Wong E., Guger K., Gumbiner B.M.;				
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.				
DR	EMBL;	U6442; AAB41671.1; -			
DR	InterPro;	IPR008938; ARM.			
DR	InterPro;	IPR000225; Armadillo.			
DR	Pfam;	PF00514; Armadillo_seg; 4.			
DR	SMART;	SM00185; ARM; 5.			
SQ	SEQUENCE 2829 AA; 310878 MW; 8A2BABDB770E496 CRC64;				

Query Match		72.4%;	Score 10544;	DB 13;	Length 2829;
Best Local Similarity		74.4%;	Pred. No. 0;		
Matches 2131;		Conservative 263;	Mismatches 413;	Indels 58;	Gaps 40;
Qy	1	MAAASVDQLLKQVEALKMENSLRQBLEDNNSHLTKLETEASNMKEVLKQLQGSIDEAM	60		
Db	1	MAAASVDQLKQVEALTKLETEATNMKEVLKQLQGSIDEAM	60		
Qy	61	ASSGQDILLERLKLNDLSNPFVGLRGMKSLRSYSGSREGSVSSRGCSPPVPMGSPFR	120		
Db	61	ASSGPDILLERFKDLNDLSNTPAGAKPKMSKSYSGSREGSLSGHSGCSPPVGVGFQR	120		
Qy	121	RGFVNSGRSTGYLELEKERSLLLDLADLKEKEKQWYVAQLNLTKRIDSLPLTNFSL	180		
Db	121	RGLLNGRSAGSYWELEKEKRLLLIAHEKEKEKQWYVAQLNLTKRIDSLPLTNFSL	180		

181 QTMTRQLEBYEARQIRVAMEBOLGTCQDMKEKQARRIARIQOIEKDIIRIQLQSQAT 240
 181 QTMTRQLEBYEARQIRVAMEBOLGTCQDMKEKQARRIARIQOIEKDIIRIQLQSQAT 240
 241 E-AERSSONKHETGSHDABEQNGQVGEINMATS--GNGGSGSTRMDHETASVLSSTH 298
 241 E-AERTPQSHDAGSDAEKLPQCGTSEITASCNVGSGQSSRADHDTTSMGSENSTY 300
 299 SAPRRLTSHGTVKVMYISLLMLGTHDKDDMSRTILLAMSSQDSCISMQRGCLPLLIQ 358
 301 SVPRRLTSHGTVKVMYISLLMLGTHDKDDMSRTILLAMSSQDSCISMQRGCLPLLIQ 360
 359 LLHGNKDSVLLGNSGSKSEARASAAALHNIHSDPDOKRGREIRVLHLEQIRAYCE 418
 361 LLHGNKDSVLLGNSGSKSEARASAAALHNIHSDPDOKRGREIRVLHLEQIRAYCE 420
 419 TCWEQEAHEPFGDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAEL 478
 421 TCWEQEAHEPFGDQDNMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAEL 480
 479 LQVDCMYGILNDHYGILNRYAGMALTNLTPGDVANKATLCSMKCMALVAQLKSESE 538
 481 LQVDCMYGILNDHYGILNRYAGMALTNLTPGDVANKATLCSMKCMALVAQLKSESE 540
 539 LQOQVIASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVLALWNLSAH 598
 541 LQOQVIASVLRNLWRADVNSKKTLEVGSVKALMECALEVKESTLKSVLALWNLSAH 600
 599 CTENKADICAVDGAFLVGTLYRSQNTWLAIESGGGILRNVSLLIATNEDHQILRE 658
 601 CTENKADICAVDGAFLVGTLYRSQNTWLAIESGGGILRNVSLLIATNEDHQILRE 660
 659 NNCLOTLLQHLKSHSLTVISNACGLWNLARSNPQDOEALWDMGAVMLKNIHSHKMI 718
 661 NNCLOTLLQHLKSHSLTVISNACGLWNLARSNPQDOEALWDMGAVMLKNIHSHKMI 720
 719 AMGSAALRNLMANRPAYKNDANIMSPGSLPSLHVRLKQALEAEADQHLSETFDNID 778
 721 AMGSAALRNLMANRPAYKNDANIMSPGSLPSLHVRLKQALEAEADQHLSETFDNID 780
 779 LSPKASHESKORHQSILGYDVFDTNRHDN--RSDNENTGNWTVLSPYLVNTVLPSSS 836
 781 LSPKTHRNKQKHQKMLCSYALDSSRHDDSCRSNGLNLTVLSPIVNTVLPSSS 840
 837 SRGSLDSSRSEKDSLERERIGLGNYPHATENPGTSSKR--GLQISTAAQIAKYMEEVS 895
 841 PRPTMDGSRPEKD---RERTAGLGNVHSTTESSGNSKRIQIQLSTT-AQISKVMEVS 895
 896 AIHTSQEDRSSGTTTELHCVTDERNALRRSSAAHNTNFTKSNENRSCMPYAKLE 955
 896 NIHLVQENRSSGASEHCHMSDERNSQRPSSNHPQSNPFTFKAESSTRGCPVAFMKE 955
 956 YKRSSNDLSNVSSSDGYSKRGQMKPISYSRDDDESKYGOVPADLAHAKHGANHMD 1015
 956 YKASNDLSNVSSSTEGYKRGQKVSYESYEDDESKYGOVPAGLAHAKHGANHMD 1015
 1016 DNGELDTPINYSKYSDEQLNGSGRSPQONERWAPKHIIEDEIKQSORQSRQOSTY 1075
 1016 DNDTELDTPIINYSKYSDEQLNGSGRSPQONERWAPKHIIEDEIKQSORQSRQOSTY 1075
 1076 PVYTESDTHLQFOPHFGQECVSPY--RSRGANGS--ETNRYVSGNHGINOVYSQSLCOED 1133
 1076 SSTENKEEKHKKFPHPFNQSENVPAITRSRGANNQVDSRVSSNLNNSKSKCHQVD 1135
 1134 DYEDDPTNYSERYSEEEQHEEF--ERPTNYSIK--YNEEKXHVDPQTDYSLKYATIPSS- 1190
 1136 DYDDDKTTPNYSERYSEEEQHEEF--ERPTNYSIK--YNEEKXHVDPQTDYSLKYATIPSS- 1195
 1191 OKQSFSSKSSQSSKTHMSSESTSTPSSNAKQNLHPSSAQSSG--OPQKAAT 1248
 1196 QKPSFPYSNNSSKQKPKKEQVSSNS--NTPTPSPNSNRQNLHPNSAQSRPGLNRPAQIPN 1254

1249 CKVSSINOETIQTVCVEDTPICFSSCSLSLSAEDEI--GCNQTTOEADSAANTLOIAEI 1307
 1255 -KPPSINOETIQTVCVEDTPICFSSCSLSLSAEDEI--GCNQTTOEADSAANTLOIAEI 1313
 1308 KEKICTGTSADPVESEPAVQHPRTKSSRLQGSLS--SESARHKAVEFSSGAKSPKSGA 1366
 1314 KE-ISAISKOGAVNETRSSVHTTKNRLQTSNISPSSDSSRHKSVEFSSGAKSPKSGA 1372
 1367 QTPKSPPEHYVQETPLMFSCSTSVSSLDSPESRSIASSVOSEPCS--CMVSGIISPDLDP 1425
 1373 QTPKSPPEHYVQETPLMFSCSTSVSSLDSPESRSIASSVOSEPCS--CMVSGIISPDLDP 1432
 1426 SPQGTMPSPSKTTPPPQTAQTKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPDA 1485
 1433 SPQGTMPSPSKTTPPPQTAQTKREVPKNAPTAEKRESGPKQAAVNAAVQVQLPDA 1489
 1486 DTLHFAFATESTPGFSCSSLSALSDELPIQKDVLRIMPVPOENDNGNETSEBPKE 1545
 1490 DTLHFAFATESTPGFSCSSLSALSDELPIQKDVLRIMPVPOENDNGNETSEBPKE 1547
 1546 NENOEKAEKTIIDSEKOLLDDDDDEIELEBECIISAMPTKSKRKA--PAQTASKLPPP 1604
 1548 IDNKAKEKRESEKQKMLDDT--DDDIDLEBECIISAMPTKSKRKA--PAQTASKLPPP 1606
 1605 VARKPSQLPVYKLLPSQNRLOPKQHSFTPGDDMPRVYCVGEPINPSTATSLDTIES 1664
 1607 VARKPSQLPVYKLLPSQNRLOPKQHSFTPGDDMPRVYCVGEPINPSTATSLDTIES 1666
 1665 PPNELAAGEVGRGAOSGEPEKDTTPTEGRSSTDEAGGKTSSVTIPELDDNKAEGDIL 1724
 1667 PPSE--PTNQDQNTDLSSTDEKDTTPTEGRSSTDEAGGKTSSVTIPELDDNKAEGDIL 1724
 1725 AECINSAMPKSKHPRVVKIMDQVOQASASSAPNKNOL--DGKKKKTPSPVKPIQNT 1783
 1725 AECIHSAMPKSKHPRVVKIMDQVOQASASSAPNKNOL--DGKKKKTPSPVKPIQNT 1784
 1784 EYRTRVRKADSKNNAERVSNDKSKQNLKNSKOFNDKLPNNEDVRGSPAFDSP 1843
 1785 GFKEERLKNTELKNPNSNQYC---DPRKPSKSKPSKANEKIPNNEETKG--FAPDSP 1840
 1844 HHTPIEGTPIYCFGRNDLSLSDPDDVDLSREKAEKREKAKENKSEAKVTSHETLSN 1903
 1841 HHTPIEGTPIYCFGRNDLSLSDPDDVDLSREKAEKREKAKENKSEAKVTSHETLSN 1900
 1904 QOSANKTQATAPINRGOKPILQKOSTPQSSKIDIPRGAATDEKLQNFALNTPVCF 1963
 1901 NPMKQDQTPKSLGGRDQPKALVQKPTSFSAAGTQDRGGATDEKMFALNTPVCF 1960
 1964 SHNSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSFHVEDTPIYCFSR 2023
 1961 SRNSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSFHVEDTPIYCFSR 2020
 2024 NSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSFHVEDTPIYCFSR 2081
 2021 NSLSLSLSDIDQENNNKENEPIKETEPDPOGEPSPQASGVAPKSFHVEDTPIYCFSR 2080
 2082 DIQPDSEHGLSPDSENFDMKATQEGANSIVSRHQAACLSRQASSDSDLSLSKSG 2141
 2081 DIQPDSEHGLSPDSENFDMKATQEGANSIVSRHQAACLSRQASSDSDLSLSKSG 2140
 2142 ISLGSPPHLTPDQBEKEFTSNKGPRLKPGKSTLETTKIESKGIKGGKVKYKSLITG 2201
 2141 ISLGSPPHLTPDQBEKEFTSNKGPRLKPGKSTLETTKIESKGIKGGKVKYKSLITG 2200
 2202 KVRNSISGOMKQPLQANMPSIRGRTMIHIQVRNNSSTSTSPVSKGPKLTPASKSP 2261
 2201 KSRSSDPSFCHCKQSVQNTNPSIRGRTMIHIQVRNNSSTSTSPVSKGPKLTPASKSP 2260
 2262 SEGQATTSRGAKPSVKSELSPVARQTSQIGSSKAPSRGSRDSTPSPQAQPSRPI 2321
 2261 NENSSSSSKPGTKP--LKSELVYGSRPSSTPGGSKGNSRSGRSDSASSRPSQPLSRPL 2319
 2322 QSPGRNISGPRNGISPPNKLSQLPRTSPSTASTSSGSKGKMSYTSPPQMQSQNLTKQ 2381

Db	2320	QSPGRNSISPGKNGISPPNKFQSPRTTSPASTKSSGSRMSYTSRGRLOSPNLSKQ	2379
Qy	2382	TGLSKNASSIPRSEASAKGLNNGNANKVELSRMSSTKSSGESDRERPVLRQS	2441
Db	2380	SGLPKTHSIPRSEASAKSLNQ-VNVTGSKKVELSRMSSTKSSGESDRERFALVRQS	2438
Qy	2442	TPIKAPPTLRKLEESASIESPSRSPASPTKSOATPVLPSPDMSLSTHSSVQA	2501
Db	2439	TPIKAPPTLRKLEESASIESSSSRADSPRSPQTQ2ALSPSLPDMALSTH-SIQA	2497
Qy	2502	GGWRKLPNLSPTIYNDGRAPKADHARSSESPLNRSCTWKBEHSHSSSLPRV	2561
Db	2498	GGWRKLPNLSPTIYNDGRAPKADHARSSESPLNRSCTWKBEHSHSSSLPRV	2555
Qy	2562	STWRRTGSSSSILSASSSESSEKAKSEDEK-HVNSISGTTQSKENQVSAKGWTRIKENEF	2620
Db	2556	STWRRTGSSSSILSASSSESSEKAKSEDEKQVCSPGPR--SECSSSAKGWTRIKESSEI	2613
Qy	2621	--SPNISTQTVSSGATNGASKTLYQMAVASKTEDVWVRIEDCPINNPRSGRPTGN	2678
Db	2614	LETSPNGSSSIATIE-SNCSLESKTLVYQMAVASKTEDVWVRIEDCPINNPRSGRPTGN	2672
Qy	2679	TPPVIDSYSSEKANPNIKDKONQAKQNGVSGVPMRTVGLNRLNSFIQVDAPQKGTET	2738
Db	2673	SPPVINDVLDQOKE-EAAKCHTEHNSGNGVPL---LENRQKSFIVDGLDTKGTDP	2727
Qy	2739	KPGQNPVPVSETNESSIVERTPPSSSSSSKHSSPSGTVAAARVTFNNVPPRKSADST	2798
Db	2728	KSLINNQ---OETNENTVAERTAFSSSSSSKHSSPSGTVAAARVTFNNVPPRKSNGENS	2784
Qy	2799	SARPSQIPTPVNNNTKRDSTDSPESSGTOSPKRHSGSYLVTSV	2843
Db	2785	TSRPSQIPTPVNTSKKRDSTETDSSGSGSPKRHSGSYLVTSV	2829

RESULT 2

Q8BNP7	PRELIMINARY;	PRT; 1056 AA.
AC	Q8BNP7	
DT	01-WAR-2003 (TRENBLrel. 23, Created)	
DT	01-WAR-2003 (TRENBLrel. 23, Last sequence update)	
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)	
DE	Adenomatosis polyposis coli (Fragment).	
GN	APC.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=Adipose tissue;	
RX	MEDLINE=22354683; PubMed=12465851;	
RA	The FANTOM Consortium,	
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;	
RT	"Analysis of the mouse transcriptome based on functional annotation of	
RT	60,770 full-length cDNAs."	
RL	Nature 420:563-573(2002).	
DR	EMBL; AK080907; BAC38073.1; --	
DR	MGI; MGI:88039; APC.	
DR	GO; GO:0005737; C:cytoplasm; IDA.	
DR	GO; GO:0005634; C:nucleus; IDA.	
DR	GO; GO:0008013; F:beta-catenin binding; IDA.	
DR	GO; GO:0009952; P:anterior/posterior pattern formation; IMP.	
DR	GO; GO:0009798; P:axis specification; IMP.	
DR	GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.	
DR	GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.	
DR	InterPro; IPR008938; Arm.	
DR	InterPro; IPR000225; Armadillo.	
DR	PIR; P00514; Armadillo_seg; 4.	
DR	PROSITE; PS0176; ARM_REPEAT; 1.	
FT	NON TER 1056 1056	
SQ	SEQUENCE 1056 AA; 117633 MW; 86C79FDC12C23FAB CRC64;	

Query Match	35.4%;	Score 5162;	DB 11;	Length 1056;
Best Local Similarity	95.1%;	Pred. No. 1-2e-245;		
Matches 1006;	Conservative 21;	Mismatches 29;	Indels 2;	Gaps 2;
Qy	1	MAAASVDQLLKQVEALKMENSRLRQELNDSNHLTKLETEASNKVKVLKOLQSGIDEAM	60	
Db	1	MAAASVDQLLKQVEALKMENSRLRQELNDSNHLTKLETEASNKVKVLKOLQSGIDEAM	60	
Qy	61	ASSGQIDLLRLKELNLDNSNPFQVKLRKMSLRSYGSRGSSVSSRGSCSPVPMGSPFR	120	
Db	61	-TSGQIDLLRLKELNLDNSNPFQVKLRKMSLRSYGSRGSSVSSRGSCSPVPMGSPFR	118	
Qy	121	RGFVNGSRSTGYLEBELEKERSLLADLDKEEKEDWYVAQLQNLTKRIDSLEPLTENFSL	180	
Db	119	RTEFVNGSRSTGYLEBELEKERSLLADLDKEEKEDWYVAQLQNLTKRIDSLEPLTENFSL	178	
Qy	181	QDTMTQRQLEVEARQIRVAMEBOLGTCDQWEKACQRIARIQOIEVDILRIROLLOSOAT	240	
Db	179	QDTMTQRQLEVEARQIRVAMEBOLGTCDQWEKACQRIARIQOIEVDILRIROLLOSOAA	238	
Qy	241	EAERSSQNHETGSHDAERQNGQGVGEINMATSGNGQSTTRMDHETASVSSSSSTHSA	300	
Db	239	EAERSSQSRHDAASHEAGRQHEGVAESNTAASSSGSQSPATRVDEHETASVSSSGTHSA	298	
Qy	301	PRELTSHLGKTVEMVYSLMLGTHDKDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL	360	
Db	299	PRELTSHLGKTVEMVYSLMLGTHDKDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL	358	
Qy	361	HGNDKDSVLLGNRSGKEARARASAAHNIHSQPDQKGRREIRVHLHLEQIRAYCETC	420	
Db	359	HGNDKDSVLLGNRSGKEARARASAAHNIHSQPDQKGRREIRVHLHLEQIRAYCETC	418	
Qy	421	WEOEAHEPQMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAIELQ	480	
Db	419	WEOEAHEQMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLQIAIELQ	478	
Qy	481	VDCEMYGLTNDHYSITLRRVAGMALNTLFGDVANKATLCSMKGCNRALVAQKSSIDL	540	
Db	479	VDCEMYGLTNDHYSITLRRVAGMALNTLFGDVANKATLCSMKGCNRALVAQKSSIDL	538	
Qy	541	QOVIASVLRNLSWRADVNSKKTLLREYGVSKALMECALEVKKESTLKSLSALWNLSAHCT	600	
Db	539	QOVIASVLRNLSWRADVNSKKTLLREYGVSKALMECALEVKKESTLKSLSALWNLSAHCT	598	
Qy	601	ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGIILRVSSSLIATNEDHQILRENN	660	
Db	599	ENKADICAVDGAFLVGLTLYRSQNTLAIIESGGIILRVSSSLIATNEDHQILRENN	658	
Qy	661	CLQTLQLKSHSLTIVSNACGTLWNLSARNPKDQALWDMGAVSMKLIHSHKHMIAM	720	
Db	659	CLQTLQLKSHSLTIVSNACGTLWNLSARNPKDQALWDMGAVSMKLIHSHKHMIAM	718	
Qy	721	GSAALRNLMANRPAYKQANIMSPGSSILPSLHVRRKQKALELDAQHLSSETPDNDLS	780	
Db	719	GSAALRNLMANRPAYKQANIMSPGSSILPSLHVRRKQKALELDAQHLSSETPDNDLS	778	
Qy	781	PKASHESKORHKOSLYGDVFDNRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSSRGS	840	
Db	779	PKASHESKORHKONLYGDYAFDNRHDDNRSDNFNTGNMTVLSPLYNTTTLVLPSSSSSRGS	838	
Qy	841	LDSSRSSEKORSLERGIGLGNHYHPATENPGTSSKRGLOISTTAAQIAKMEVSAIHTS	900	
Db	839	LDSSRSSEKORSLERGIGLSAYHPTTENAGTSSKRGLOITTTAAQIAKMEVSAIHTS	898	
Qy	901	QEDRSSGSGTTELHCVTDERNALRSSAAHTHNTYNTFTKSENSNRCSPYAKLEYKRS	960	
Db	899	QDDRSSASITTEPHCVADRRSASHTHNTYNTFTKSENSNRCSPYAKLEYKRS	958	
Qy	961	NDSLNSVSSSDGYGKRGKQMKPSIESYSEDESKFCYGOYPADLAHKIHSANHVDNDGE	1020	
Db	959	NDSLNSVTSDDGYGKRGKQMKPSIESYSEDESKFCYGOYPADLAHKIHSANHVDNDGE	1018	

QY 1021 LDTPINYSKYDEQLNSGRSPSONERWARPKHIED 1058
DB 1019 LDTPINYSKYDEQLNSGRSPSONERWARPKHIED 1056

RESULT 3

Q921K7 ID Q921K7 PRELIMINARY; PRT; 2274 AA.
AC Q921K7;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE APC2 protein.
GN APC2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
Miles A., Kuipers J., Destree O., Feifer M., Clevers H.,
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
tumour suppressor";
RL Curr. Biol. 9:105-108 (1999).
DR EMBL; AJ130783; CAAL0207.1; JOINED.
DR EMBL; AJ130784; CAAL0207.1; JOINED.
DR EMBL; AJ130785; CAAL0207.1; JOINED.
DR EMBL; AJ130786; CAAL0207.1; JOINED.
DR EMBL; AJ130787; CAAL0207.1; JOINED.
DR EMBL; AJ130788; CAAL0207.1; JOINED.
DR EMBL; AJ130789; CAAL0207.1; JOINED.
DR EMBL; AJ130790; CAAL0207.1; JOINED.
DR EMBL; AJ130791; CAAL0207.1; JOINED.
DR EMBL; AJ130792; CAAL0207.1; JOINED.
DR EMBL; AJ130793; CAAL0207.1; JOINED.
DR EMBL; AJ130794; CAAL0207.1; JOINED.
DR EMBL; AJ130795; CAAL0207.1; JOINED.
DR EMBL; AJ130796; CAAL0207.1; JOINED.
DR PIR; T30258; T30258.
DR MGD; MGI:1346052; Rnc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00514; Armadillo_seg_3.
DR SMART; SM00185; ARM; 5.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
SQ SEQUENCE 2274 AA; 243137 MW; 75ABDA15D0F707F5 CRC64;

Query Match 24.1%; Score 3512.5; DB 11; Length 2274;
Best Local Similarity 34.3%; Pred. No. 4.9e-164;
Matches 1005; Conservative 354; Mismatches 788; Indels 779; Gaps 97;

QY 4 ASVDOLKQVEALKWENSLRQELNLSNLTLEASNMKEVLKQLOCSIDEA--MA 61
DB 6 ASVEQLVRQVEALKWENSLRQELNLSNLTLEASNMKEVLKQLOCSIDEA--MA 65
QY 62 SSGQIDLLERKLEINLDSNPPGVKRSKMSLSYSGREGSVSRSGECSVPVPMGFPFR 121
DB 66 SSGQTEVLEQLKALQTDISSLYNLKHPA---ALGPEP---AARTPEGSPV-HGSGPSPK 117
QY 122 -GFVNGRSTGYLEERLEKERSLLADLKEEKDQWYLAQLNLTAKRIDSLPLTENFSL 180
DB 118 DSFGELSRATIRLEEDLQRCFLLSIEKEKEKLWYSLQGLSKRLDELPHVDTFSM 177
QY 181 QDTMTREQLYEAFQIRVAMEQLGTQDMKEKPAQRRIAFIQIEXDIL----RIQ--- 233
DB 178 QMDLIRQLEFEAGHRSLEMEERFGTSDMVQRAQIRASLEQIDKELLEAQDRVQOTEP 237
QY 234 --LLQSQATERSRSQNKHETGSHDAERQEGQVGEINMATSGNGSGTTRMDHETASV 291
DB 238 QALLAVRPVAVEEQEAETVHPDGTGPQ-----PQN----- 269

QY 292 LSSSTHSAPRRLTSHLGTKEVYVYLLMLGTHDKDMRSTLLAMSSSDSCISMRSQSG 351
DB 270 -----SKVEVVFLLSLMFLATDQEDTARTLLAMSSSPESCVAMERSG 311
QY 352 CLPLLIQLLHGNKDSV---LLGNRSRSGKEARAPASAAHLNIIHSDPDDKRGREIRVLH 408
DB 312 CLPLLIQLLHGNKDSV---LLGNRSRSGKEARAPASAAHLNIIHSDPDDKRGREIRVLH 408
QY 409 LLEQIRAYCETCEWQSAEPGMDQKNPMPAPVEHQICPAVCVLMKLSDEDEHRAVME 468
DB 372 VLEQIRAYCETCEWQSAEPGMDQKNPMPAPVEHQICPAVCVLMKLSDEDEHRAVME 468
QY 469 LGGQIAELQVDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKMGQMA 528
DB 427 LGGQIAELQVDCENYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCMKMGQMA 528
QY 529 LVAQLKSESEDLOQVIAVLNLSWRADVNSKTLREVGSKVLMKALMECALBVKESITLKV 588
DB 487 IVAQLGSESEELHQVSSILNLSWRADVNSKTLREVGSKVLMKALMECALBVKESITLKV 588
QY 589 LSALWNLSAHCTENKADICAVDGAFLVGLTLVRSQTNLTALIESGGGILRVSSLIAT 648
DB 547 LSALWNLSAHCTENKADICAVDGAFLVGLTLVRSQTNLTALIESGGGILRVSSLIAT 648
QY 649 NEDHROILRENNCLQTLLOHLKSHSLTIVSNAGCTLWNLARPQKOEALWDMGAYMLK 708
DB 607 REDYRQVLRDHNCLQTLLOHLKSHSLTIVSNAGCTLWNLARPQKOEALWDMGAYMLK 708
QY 709 NLIHSHKMTAMGSAALRNLMANRPKYKDAMT-MSPGSSLSPLSHVRKQKALEAELDAQ 767
DB 667 NLVESKHKMTAMGSAALRNLMANRPKYKDAMT-MSPGSSLSPLSHVRKQKALEAELDAQ 767
QY 768 HLSETFDNIDNLS-PKASHRSKQ-----RHQKSLYGDYVFTDTHRDDNRSNFTGNT- 820
DB 727 HLVAHGLHLEKQSLPEAETTSKKPLPLRLHDLGLVQYASDSGCFDDDDAFSAAATA 786
QY 821 -----VLSPLYNTVLPSSSSRSGSLDSRSEKDRSLERERGIGLGNYPHATENPTSS 874
DB 787 EPASPAVNSMFLGGPFLQGGALAR-----TPPARQGGLEAKAG-----GEAAVAA 833
QY 875 KRGQLISTTAQIAKAMEEVSIAHTSQEDRSSGTTTLCVTDTERNALRRSSAAHTSNT 934
DB 834 KAKALALAVARIDRLVEDISALHTSSDDGFSLS-----GDPQAPREGRAQSCSPC 887
QY 935 YNFT--KSENSNRCSMPYAKLEYKRNSDLSNLSVSSDYGKGGKQKPSIESYSEDES 992
DB 888 RGTEGGRREGSRAHPLRLKAAHTSLNSDLSNLSGSGTSDGYCTREHMT- 936
QY 993 KFCYGVQPADLAHAKIHSANHMDDNDGELDTPINYSKYDEQLNSGRSPSONERWARP 1052
DB 937 --CLPAL-----AEHRDD-----PVRQ--TRP 956
QY 1053 KHIDEIKQEQEQSQNRQSTTPYVYTESDTHKLPQHPGQEQCVSPRSRGANGSET 1112
DB 957 RRLDILPSRAELPARDTAATDARVT-----IKLSPTYQHVPFLD-----GAAGA-- 1002
QY 1113 NRVGSHGINQVSSQLCEDDYDDKPTNYERYSEEEQHEEERPTNYSIKYNEEKRH 1172
DB 1003 -----GVRPLVPGTGS-----AEHRDD-----PVRQ--TRP 956
QY 1173 VDQPIDYSLKATDIPSSQKQSFSSKSSQSGSKTEHMSSSSTSTPSSNAKRONQLH 1232
DB 1014 -----PCARKQAW----- 1021
QY 1233 PSSAQRSGQPKAATCKVSSINQETIQTVCVEDTPICFSRCSLSSLSAEDEIGCNOT 1292
DB 1022 -IPADLSKVPEKLVASPL-PIASKVQKLVQAGPNSLRCSSLSLSSTGHAVPSQAE 1079
QY 1293 TOEADSAANTLOIAIEKIGITRSAB-----DPVSEVPASVQHPTKSRLOGS 1340
DB 1080 NLDSDS-----LEGLEAGPGEALGAMRASGTSILPVS-IPA-----PQGRSR----- 1125

QY 1341 SLSSSARHKAVERFSSGAKSPKSGAQTQPKPPEHYVOETPLMFSRCTSVSSLSDFERS 1400
DB 1126 GLGVEDA-----TFSSSENVCVQETPLVLSRCSVSSLSGFSERS 1165
QY 1401 IASSVQSPBCSMGWSIIPSDLPSPGOTMPPSRKTPPPPPQTAQTKR-----EVPK 1454
DB 1166 IASSIPSPBCSLGSGTVPBELPSPGOTMPPSRKTPPPPPQTPETQSLOWESYVK 1225
QY 1455 NKAPTAEKREGPKQAANAQVQVLPDADTLHFFATEPTDGFCSLSLSALSLDEP 1514
DB 1226 RFLDIADCRERCQPPSELDAGSVR-----FTVEKPDENFSCASSLSALHEL 1273
QY 1515 FIOKDVEIRLIPPV-QENDNGNEIETESEQPKSENQKEAKTIDSEK---DLDDSDDD 1570
DB 1274 YVQDVEIRLIPVPPACPERAVG-----GHRHRRDEAASRLDPAPAGRSARSATDK 1324
QY 1571 DIEILEECIIISAMPTKSKRKAQPAQATKSLPPVPVARKPSQLPVYKLPSQNRLOPKHV 1630
DB 1325 ELEALRECLGAAMPARLAK-----VASALVP--GRSLPVPVTVMLVPAPAR----- 1368
QY 1631 SFTPGDDMPVYCVGEGTINSTATSLDILIESPPNELAAGEVURGGAQSGFEKEDTI 1690
DB 1369 ---GDD-SGTDSABGTVPNFSSAASLSDTLQGPSRDKPAGPGR-----OK----- 1411
QY 1691 PTEGRSTDEAOGKTSVTIPELODNKAEGDIIAECINSAMPKGSHPKPRVKIMQV 1750
DB 1412 PT-GEAARQO-----TRSRP----- 1427
QY 1751 QOASASSAPNKLQDKKKKPTSPVKI PONTY-RTVRKNADSKNNLNAEVRFSNKK 1809
DB 1428 KAAGAGKS-----TEHTRGFCNRRAGLEPLSRPOSARSNR 1463
QY 1810 DSKQNLKNNKDFNDKPNNEDEVRGSPAFDPSPHYTPIEGTPVCFSRNDSLSLDFDD 1869
DB 1464 DSSCOT-----RTRGGLAQSLCLTTPTEAVYCF-----YDS 1496
QY 1870 DDVULS-----REKAELRKAKENKESBAKYTSHTELTSNQQSANKTOAIKQPIN--R 1920
DB 1497 DEEPPATAPPERRAGAIPLALKREKPAKX-----ETPSRAAQATLIPVR 1541
QY 1921 GQPKILOQSTFFPOSSKDI PDGNAATDEKLNFAIENTPVCFSHNSLSLSLSDIDENN 1980
DB 1542 AQPLRI-----VDTPCYSLTSSASLSB----- 1566
QY 1981 NKNEPIKETBPDDQSGFSPQAGYAPKGFHVHEDTPVCFSRNDSLSLSDIDEDLLQ 2040
DB 1567 -----PEA---PEQPANHARGPEQSKOD-----SSPSPRAEEELQ 1600
QY 2041 ECISSAMP-----KKKPSRLKGNKXSPNMGILGEDTLDLKDIQRPDSR-H 2090
DB 1601 RCISLAMPRRRTQVFGSRRRKRPLRSD-----IRPTEIT-----QRCQBEVA 1643
QY 2091 GLSP--DSENFQKAIQEGANSIVSSLHQAACLSRQASSDSLSLSKSGISLSPF 2148
DB 1644 GSDPASDLSVQWQAIQEGANSIVTLHQAAKASL--EASESDLSLSVSGVSAGSTL 1701
QY 2149 HLTPDQBEKPTNSKGRPILKPEKSTLETXXI-ESESKIGKGGKVKVYKSLITQVR-SN 2206
DB 1702 Q-----PSKLARKGRKPAABAGAWRPEKRGTTSTK-----INGSRPLPN 1740
QY 2207 SEISQMKCPLQANPISIRGRTMHIHGVNNSSTSPVSKGPPPLKTPASKSPSGQT 2266
DB 1741 GPEKAGTQKQWAGSTMLRGTVIY-----SAGPASTQSKG--ISGPTTTPKKTGTS 1792
QY 2267 ATTSPRGAKPVKSELSVPARQTSOIGGSKAPGRSGRSDTPRPAQPLSRPIQSPGR 2326
DB 1793 GTTQPETV-----TKAPSPQQRSRLRHPGKISELALRHPPR 1831
QY 2327 NSISPRNGISPPNKLSQLPRTSSP-----STATKSSG-----SKMYSYTFGQMGOQN 2377
DB 1832 SATPPARLAKTPSSSSQTSFASQPLPRRSLPATPTGFLPCPGGSLVPKSPARAL----- 1897
QY 2378 LTKQTGLSKNASSI-----PRSESASKGLNQMNNGCA-NKKVELSEMS 2420

RESULT 4

O95996 PRELIMINARY; PRT; 2303 AA.
ID O95996 AC O95996; DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE APCL protein.
GN APCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Koyama K., Nakagawa H., Nakamura Y.;
RT "APCL exon14.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99040663; PubMed=9823329;
RA Nakagawa H., Murata Y., Koyama K., Fujiyama A., Miyoshi Y., Monden M.,
RA Akiyama T., Nakamura Y.;
RT "Identification of a brain-specific APC homologue, APCL, and its
RT interaction with beta-catenin.";
RL Cancer Res. 58:5176-5181 (1998).
DR EMBL; AB022529; BAA75469.1;
DR EMBL; AB022518; BAA75469.1; JOINED.
DR EMBL; AB022519; BAA75469.1; JOINED.
DR EMBL; AB022520; BAA75469.1; JOINED.
DR EMBL; AB022521; BAA75469.1; JOINED.
DR EMBL; AB022522; BAA75469.1; JOINED.
DR EMBL; AB022523; BAA75469.1; JOINED.
DR EMBL; AB022524; BAA75469.1; JOINED.
DR EMBL; AB022525; BAA75469.1; JOINED.
DR EMBL; AB022526; BAA75469.1; JOINED.
DR EMBL; AB022527; BAA75469.1; JOINED.
DR EMBL; AB022528; BAA75469.1; JOINED.
DR EMBL; AB012162; BAA34611.1;
DR HSSP; Q02248; 3BCT.

DR	GO; GO:0008013; F:beta-catenin binding; TAS.	DB	799	PAULSLFLGSPFLQGGQALARTPTTRSG--GKEAEKDTSGE-----AAVA 840
DR	GO; GO:0006461; P:protein complex assembly; TAS.	QY	874	SKRGLQISTTAQIAKIMVEVSALHTSQEDRSGSGSTTEHLCHVTDERNALRRSAAHTHS- 932
DR	GO; GO:0007165; P:signal transduction; TAS.	DB	841	AKAKAKALAVARIDQLVEDISALHTSSDDSFSLSS-----GDPQGEAPREGRQAQSCSP 894
DR	InterPro; IPR008938; ARM.	QY	933	NTYNFTKSNRNRTCSMPYAKLVEYKRSNDSLNSVSSDGYGKQKQMKPSTESYSEDDE 991
DR	InterPro; IPR001225; Armadillo.	DB	895	CRGEGGRRRAGSGRAHPLLELKAHAASLNSLNSGASDGYCPREHM----- 942
DR	pfam; PF00514; Armadillo_5eg; 3.	QY	992	SKFCSYGQYPADLAHKTHSANHVDNDGELDTPTNYSLKYSEQLNSGQSPQSONERWAR 1051
DR	SMART; SM00185; ARM; 5.	DB	943	LPCLPLA-----ALASREDF----- 957
DR	PROSITE; PS00546; CYSTEINE_SWITCH; 1.	QY	1052	PKHIIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHLKFQPHFGQBCVSPYRSRGANGSE 1111
DR	SEQUENCE 2303 AA; 243946 MW; 7BF940183ACD643D CRC64;	DB	958	RCQGPRESR----- 966
DR	Query Match 23.4%; Score 3415.5; DB 4; Length 2303;	QY	1112	TNRVGSNHGINQVNSQSLCQEDDYEKKFTNYISERYSEBEQHEEERPTNYIKYNEEKR 1177
DR	Best Local Similarity 34.4%; Pred. No. 3e-159;	DB	967	LDLPLCQCAEPPAREATSADARVRI---IKLSPTYQ 1000
DR	Matches 997; Conservative 372; Mismatches 815; Indels 717; Gaps 94;	QY	1172	HVDQPIYSUKYATDTPSSQKQSFSSKSSGQSKTEHMSSESSENTSTPSSNAKQNL 1231
DR		DB	1001	PLLEGASRAGAEP LAGPGI-----SPGAKQAWL 1031
DR		QY	1232	HPSSAQSRSGOPOKAATCKVSSINQETIQYCVEDTPICFSRCSLSSLSLSAEDIGCNG 1291
DR		DB	1032	PADHLKSVPEKLAAPL-SVASKALQKLAQEGFLSKRCSLSSLSLSA-GRPGPSE 1086
DR		QY	1292	TTQPADSANTLOIAEIKKEKIGTRSAR-----DPVSEVPVAVSQHPRTKSKSLQSSLS 1344
DR		DB	1087	GGDLDDSDSSL---EGLEEAGPSEALDSTWRAPGATSLPVAIPAPR-----RNRGRGLGV 1139
DR		QY	1345	ESARHKAVERSSGNAKSPSKSGAOTPKSPPHVYQETPLMFSRCTSVSSLDSPESRSIASS 1404
DR		DB	1140	EDATPSSSENHYQETPLVLSRCSVSSLSGTSFSPSIASS 1179
DR		QY	1405	VQSEFPCSMVSGIISPSDLDPSPQTMPPRSRSTPP--PPQTAQTKREVVPKXKAAPTAEK 1462
DR		DB	1180	IPSEPCSGQSGITSPSELFDSPQTMPPRSRSTPPLAPAPQ-----PPEATQFSLQ 1232
DR		QY	1463	RESGPKQAVNAVQRVQVLP---DADTLHLHATESTPDGFCSSSLSALSIDEPFQKD 1519
DR		DB	1233	WESTVYKFLDIADCRERCRLPSELDAQS-VRFVTEKPDENFSCASSALALHHEHYOOD 1291
DR		QY	1520	VELRIMPVQVENDNGNTESEQPKSENENQEKAEKTIIDSEKDLDDSDDDIEILEECI 1579
DR		DB	1292	VELRLPLSACPERGGAGGAGLHFAHRRREEGPAFTGSRPR---GAADQELLLRECL 1347
DR		QY	1580	ISAMPTKSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPSQNRQLPQKHVSTFGDDMP 1639
DR		DB	1348	GAAPARLRK-----VASALVP--GRRALPVVYMLVPAPAPQ-----EDDS 1388
DR		QY	1640	RVCYVEGTPINFSTATSLDLTIESPNELAAAGEVGGAGSQSEFKEKRTIPITEGRSTDE 1699
DR		DB	1389	CTDSAEGTPVNFVNSAASLDELTLGGPRDQPGGAPR-----QRTT---GRPTSA 1435
DR		QY	1700	AQGGKTSSTVPIPELDDNKAEEGDILAEICINSAMPKGSHKPRVKKIMDQVOQASASSA 1759
DR		DB	1436	RQA-----MGRHK-----AGGAGS 1451
DR		QY	1760	PNKNQLDGKK-----KKPTSPVKPIPQNTYETRVKKNADSKNLNAERVSFNDKD 1810
DR		DB	1452	AEQSRGAGKQNRAGLEPLGRPPSAPA-----DKD 1480
DR		QY	1811	SKQNLKNNSKDFKDLNPNNEDEVRSFAPDSPHHYTPIEGTPCYFSRNDLSLSDFFDD 1870
DR		DB	1481	GSKG-----RTRGDGALQSLCLTTPTEEAVCYFGNDS----- 1514
DR		QY	1871	DVDSRKAELRKAKENKEAKVTSHTELTSSNQOQANKTQAIKQPINRGQPKPILOKQ 1930
DR		DB	1515	DEEPPAAAPTTH-----RETSAIPR-AFTREPR---QGR 1543

DE	APC2 protein (Fragment).	
GN	APC2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Kidney;	
RX	MEDLINE=99147086; PubMed=10021369;	
RA	van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,	
RA	Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;	
RT	"Adenomatous Polyposis Coli Homologs in Mammals and Flies.;"	
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AJ012652; CAB61207.1;	
DR	EMBL; AF128222; AAF01784.1;	
DR	InterPro; IPR008938; ARM.	
DR	InterPro; IPR000225; Armadillo.	
DR	Pfam; PF00514; Armadillo_seg; 3.	
DR	SMART; SM00185; ARM; 5.	
FT	NON TER 733	
FT	SEQUENCE 733 AA; 80876 MW; 09B56B5EF7032BAD CRC64;	
SEQ		
Query Match	14.9%; Score 2170.5; DB 4; Length 733;	
Best Local Similarity	57.5%; Pred. No. 1.2e-98;	
Matches	445; Conservative 122; Mismatches 154; Indels 53; Gaps 11;	
QY	2 AAASYDQLLKVEALKVENSLRQLEDNSHLTKLETASNMKEVLKOLQGSIEDA-- 59	
DB	4 SVAPYEQVLRQVQKALKAENSHLRQELDRNSHLSKLETETSGMKEVLKLGKLEQEARV 63	
QY	60 MASSGQIDLLERLKLKVENSLRQLEDNSHLTKLETASNMKEVLKOLQGSIEDA-- 119	
DB	64 LVSSGQTEVLEQLKALQMDITSLYNLKFQPP---TLGPEP---AARTPEGSPV-HGSGP 115	
QY	120 RR-GFVNGSRESTGYLBELEKERSLLADLDKEKEDWYVAQIQLTKRIDSPLTFNF 178	
DB	116 SKDSFGELSRATIRLLELDRECRFLNEIEKEKEKLMWYSQGLSKRLDELPHVTF 175	
QY	179 SLQDTMFRQLEYEARQIRVAMEEQLTGCDMEKRAQRRIARIQIEKDIRIRQLLSQ 238	
DB	176 SMQVLLIRQLLEFPAQHRSILMEERFCTSDMVQRAIRASRLQIDKELLE----- 227	
QY	239 ATEARSSQNHETGSHDAERQEGQGVGEINMATSGNGQSTTRMDHETASVLSSSTH 298	
DB	228 -----AQDRVQOTEPPQALLAV-----KSPVDEDPETEVPHTPED 262	
QY	299 SAPRLTSLGTQVEMVYSLLSMLGTHDKDDMSRTLLAMSSSDSCISMRQSGCLPLLIQ 358	
DB	263 GTPQPGNS-----KVEVFWLLSLMTRDQEDTARTLLAMSSSPESCVAMRSCLPLLIQ 318	
QY	359 LLHGNDKDS---VLLGNSRGSKEARAPASAAALNHIHSQDDKRGRRIRVHLHLEQIRA 415	
DB	319 ILHGTAAAGGRAGAPGAPGAKDARMEANAALHNIVFSQDQGLARKEMEVHLVLEQIRA 378	
QY	416 YCTCMEWQEAHBEQMD-QDKNPPAPVEHOICPANCVLMLKLSFDEEHHANMELGLQA 474	
DB	379 YCETCDWQLQARDGGPGGGAGSAPIPEIQICQATCAVMKLSFDEEYRRANMELGLOA 438	
QY	475 IAEQLQVDCMYGLTNDHYITLRRYAGMALTNLTFGDVANKATLCMKGCRALVAQLK 534	
DB	439 VAEQLQVDYEMHKWTRDPLNLRARYAGMTLTNLTFGDVANKATLCARRGCMBAIQA 498	
QY	535 SESEDLQOVTASVLBNLSWADVNSKTLREVSQVAMCALEVKESTKLSALWN 594	
DB	499 SDSEELHQVSSILRNLSWEADINRSKVLREAGSVTALVQCVRATKESTKLSVLSALWN 558	
QY	595 LSAHCTENKADICAVDGLAPLVGTLTYRSQTNLTALIESGGGILRNVSLLIATNEHQ 654	
DB	559 LSAHSTANKAAICQVDGALGFLVTLTYKQCSNLAIIESGGGILRNVSLLVATREYRQ 618	
QY	655 ILRENNCLQTLLOHLKSHSLTIVSNACGTLWNLSARNPKDQEAALWDMGAVSMKLNLIHSHK 714	
DB	619 VLDRHNCQLTLLOHLKSHSLTIVSNACGTLWNLSARSARDQELLWDLGAVGMLRNLAHSHK 678	
QY	715 HKMIAMGSAALRNLMANRPAYK-DANIMSPGSLPSLHVVRKOKALEDAQ 767	
DB	679 HKMIAMGSAALRNLLAHPAKHQAAATAVSPGSCVPSLYVRKORALEALDPE 732	
RESULT 7		
ID	Q9Y632 PRELIMINARY; PRT; 1246 AA.	
AC	Q9Y632;	
DT	01-NOV-1999 (TReMBLrel. 12, Created)	
DT	01-NOV-1999 (TReMBLrel. 12, Last sequence update)	
DT	01-OCT-2003 (TReMBLrel. 25, Last annotation update)	
DE	APC2 protein (Fragment).	
GN	APC2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Carf i.M., Markham A.F., Colleta P.L., Wai L., Askham J., Morrison E.,	
RA	Meredit D.M.;	
RT	"APC2 alternatively spliced cDNA sequence.;"	
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF110334; AAD28183.1;	
DR	HSSP; Q02248; 3BCT.	
DR	GO; GO:0005578; C:extracellular matrix; IEA.	
DR	GO; GO:0004222; F:metalloendopeptidase activity; IEA.	
DR	GO; GO:0006508; F:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR008938; ARM.	
DR	InterPro; IPR000225; Armadillo.	
DR	InterPro; IPR001818; Pept M10A_M12B.	
DR	Pfam; PF00514; Armadillo_seg; 3.	
DR	SMART; SM00185; ARM; 3.	
DR	PROSITE; PS00546; CYSTEINE_SWITCH; 1.	
DR	NON TER 1246	
FT	SEQUENCE 1246 AA; 132115 MW; 8C527DA2B3B0BC17 CRC64;	
SEQ		
Query Match	13.7%; Score 1991.5; DB 4; Length 1246;	
Best Local Similarity	30.5%; Pred. No. 1.5e-89;	
Matches	579; Conservative 193; Mismatches 428; Indels 697; Gaps 41;	
QY	2 AAASYDQLLKVEALKVENSLRQLEDNSHLTKLETASNMKEVLKOLQGSIEDA-- 59	
DB	4 SVAPYEQVLRQVQKALKAENSHLRQELDRNSHLSKLETETSGMKEVLKLGKLEQEARV 63	
QY	60 MASSGQIDLLERLKLKVENSLRQLEDNSHLTKLETASNMKEVLKOLQGSIEDA-- 119	
DB	64 LVSSGQTEVLEQLKALQMDITSLYNLKFQPP---TLGPEP---AARTPEGSPV-HGSGP 115	
QY	120 RR-GFVNGSRESTGYLBELEKERSLLADLDKEKEDWYVAQIQLTKRIDSPLTFNF 178	
DB	116 SKDSFGELSRATIRLLELDRECRFLNEIEKEKEKLMWYSQGLSKRLD----- 167	
QY	179 SLQDTMFRQLEYEARQIRVAMEEQLTGCDMEKRAQRRIARIQIEKDIRIRQLLSQ 238	
DB	168 ----- 167	
QY	239 ATEARSSQNHETGSHDAERQEGQGVGEINMATSGNGQSTTRMDHETASVLSSSTH 298	
DB	168 ----- 167	
QY	299 SAPRLTSLGTQVEMVYSLLSMLGTHDKDDMSRTLLAMSSSDSCISMRQSGCLPLLIQ 358	

Db 168 ----- 167
QY 359 LLHGNDKDSVLLNGSRGSEARASAAALHNIHSQDDKRGRRIRVHLHLQRAYCE 418
Db 168 ----- 167
QY 419 TCWEQEAHEPGWDQKNMPAPVEHICPAVCVLMKLSFDBEHRHANVELGLQAI AEL 478
Db 168 ----- EL 169
QY 479 LOVDCMYGLTNDHYSITLRRVAGMALNLTGDEVANKATLCSMKGCMRALVAQKSSSE 538
Db 170 LOVDYEMHKTWTRDPLNLAL-RVAGMTLNTLTGDEVANKATLCARRGCM EAI VAQLASDE 228
QY 539 DLOQVTLASVLRNLNRADVNSKTLREVSGVKALMECALEVKKESTLKSVALNLSAH 598
Db 229 ELHQVSSILRLNLRADVNSKTLREVSGVKALMECALEVKKESTLKSVALNLSAH 288
QY 599 CTENKADICAVDGAALFVGLTYTSQNTLAIIESGGGILRNVSLLIATNEDHQILRE 658
Db 289 STENKAAICQVDGALGLFVSTLYTKQCSNLAIIESGGGILRNVSLLVATREDYQVLRD 348
QY 659 NNCLQTLQHLKASHSLTIVSNACGTLNLSAENPKDQALWDMGAVSKMLIHSKHOMI 718
Db 349 HNCLOTLLQHLTSHSLTIVSNACGTLNLSAENPKDQALWDMGAVSKMLIHSKHOMI 408
QY 719 AMGSAALNLMANRPAYK-DANIMSPGSSLPVSLHVRKQKALEAELDAQHLSETFDNID 777
Db 409 AMGSAALNLMANRPAYK-DANIMSPGSSLPVSLHVRKQKALEAELDAQHLSETFDNID 468
QY 778 NLSPKASHRSKO-----RHKOSLYGDIVFDNTHDDNRSNF-----NTGNMTVLS 823
Db 469 KGPFAAEAAATKPLPPLRLHLGLAQDYASDSCGDFDDAPSSLAATAATGEPASPAALS 528
QY 824 PYLNTVLPSSSSSR---GSLDSSSEKDRSLERERIGLGNHYHPATENPGTSSKRGLOI 880
Db 529 LFLGSPFFLQQAQATPTTRCGKEAKDTSG-----AAVAKAKAKL 572
QY 881 STTAQIAKWEVEIAHTSQEDRSSGTTLHCVTDERNALRSSAAHTHS--NTNYFT 938
Db 573 ALAVARIDQLVEDISALHTSSDSDSFLSS-----GDPQZAPREGRAQSCSPCKGPEGG 626
QY 939 KSESNRITCSMPYAKLEYKRSNDSLSVSSSDGYGKGOMKPSIESYDEDESKFCSYG 998
Db 627 RREAGRAHPLRLKAAHASJNDSLSGASDGYCPREHM----- 667
QY 999 QYPADLAHKIHSANHMDDNGELDTPINYSKLYSDEQLNSGRQSPSQNERWARKHIIE 1058
Db 668 -LPCPLA-----ALASREDP----- 682
QY 1059 EIKQSEQRNQSTTYPVYTESTDDKHLKFPQHPGQECVSPYRSGANGSETRVGSN 1118
Db 683 -----RCGQPRPSR----- 691
QY 1119 HGINQWSQLCQEDDDKPTNYSERYSEEQHEEERTNYSIKYNEEKRHVDQPID 1178
Db 692 -----LQDLPFGCAEPAREATSADARVT---IKLSPYQHV----- 727
QY 1179 YSLKYATDIPSSQKQSFSSKSSQSKTEHMSSSSENTTPPSNAKRONQLHPSSAQSS 1238
Db 728 -----PLLEGASRACAPLAGFI-----SPGARKQAWL---PADH 760
QY 1239 RSGQPKAATCKVSSINOETITQYCVETDTPICFRCSLSLSAEDDEIGCNOITQEADS 1298
Db 761 LSKVPEKLAAPL-SVASKALQKLAQEGPLSLGRCSLSLSA-GRPGESEGDLDSS 818
QY 1299 ANTLQIAE-----IKEKIGTSAEDPVSEVPVAVSQHPRTKSSRLQSGSSLSSEARHKAVEF 1354
Db 819 DSSLGLEEAGFIEALDSTWRAPGATSLPVAIPAPR---RNRGRGLGVEDA----- 867
QY 1355 SSGAKSPKSAQTPPEHYVQETPLMFRCTSVSSLDSPESRSIASSVQSPFCGMV 1414
Db 868 -----TPSSSENHYVQETPLVLSRCSVSSLGFSFSPSIASSIPSEPCSGH 914

QY 1415 SGIITSPDLPDPSQOTMPPSRSKTPP--PPQTAQTKREVFPKNKAPTAARESGPKQAAV 1472
Db 915 SGTISPSLSPDPSQOTMPPSRSKTPP LAPAQG-----PPEATQPSLQWESVYKRF 967
QY 1473 NAAVQRVQVLP---DADTLHTEATEPDCGSSSLSALSJLDEPFTOKDVELRIMPVQ 1529
Db 968 IADCRERCRIPSELDAAGS-VRFVTERPDENFSCASSLSALALHEHYVQDVELRLLPSAC 1026
QY 1530 ENDNGNTESEQPKESNENOEKAEKTIIDSEKOLLDDDDDDIIEIEECIISAMPTKSSR 1589
Db 1027 PERGGAGGAGLHFAHRRRGAGAGFSPSR-----RRGOELLERECLGAAPARLRK 1081
QY 1590 KAKPAQATASKLPPPVARKPSQLPVYKLLPSONRLQPKHVSFTPGDDMPRVVCVEGTPI 1649
Db 1082 -----VASGLVP---GRRALPVVYMLVPAPAPQ-----EDDSCDTSAEGETPV 1122
QY 1650 NFSTATSLDLTIESPPNELAAGEVGRGAQSGEFKRDITPTEGRST-----DEAQQGKT 1705
Db 1123 NFSSAASLSDETLOGPRDQPGGAGR-----QRPTRPTSAQAMGHRHKAGAGR 1174
QY 1706 SSVTIPELDDNKAEBGDILAEICINSAMPKSHKPRVKKIMQVQOASASSAPNKNQL 1765
Db 1175 SABQSPGAGKXRA-----GLEPLG----- 1194
QY 1766 DGKKKPTSPVKPIPQNTVEYTRVRKNADSKNNLNAERVSFDMKSKQKQKNNKSKDFND 1825
Db 1195 -----RPPSAA-----DKDSKPG----- 1209
QY 1826 KLPNNEDRVGSAFOPPHHYTFIEGTPYCFSRNDSL 1862
Db 1210 -----RTRGDGALQSLCLTPTTEAVCYGNDVS 1239
RESULT 8
P91667 PRELIMINARY; PRF; 2416 AA.
ID P91667 AC P91667;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE ADENOMATOUS polyposis COLI.
EN APC OR D-APC OR CG1451.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=37144426; PubMed=8990193;
RA Hayashi S., Rubinfeld B., Souza B., Polakis P., Wieschaus E.,
RA Levine A.J.;
RT "A Drosophila homolog of the tumor suppressor gene adenomatous
RT polyposis coli down-regulates beta-catenin but its zygotic expression
RT is not essential for the regulation of Armadillo.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:242-247(1997).
DR EMBL; U77947; AAB41404.1; -;
DR PIR; T13825; T13825.
DR FlyBase; FBgn0015589; Apc.
DR GO; GO:0008013; F-beta-catenin binding; IDA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg. 5.
DR SMART; SMO0185; ARM; 5.
DR PROSITE; PS0176; ARM_REPEAT; 2.
SQ SEQUENCE 2416 AA; 261405 MW; 4DE3A10BE10B42A9 CRC64;
Query Match 12.2%; Score 1780.5; DB 5; Length 2416;
Best Local Similarity 25.4%; Pred. No. 9.le-79;
Matches 720; Conservative 329; Mismatches 847; Indels 935; Gaps 100;
QY 243 ERSQNKHETGSHDAERQN-EGQGVGEINWATSGNQSGTTR-----MDHETASVLSSS 295

129 ELREMEHRSLDNFRQSAQQOOLDELPRNGGGSPASAGRPSSRSPSYLSRFLGDD 188
 296 STHSAPR-----RLTSH-----LGTQVEMVYSLLSMLGTHDKDDMSRTLLAM 337
 189 APAPARLPKGAAWTTSFERYTSSAVEATLGSKEVCYVLSLMLGNDPLEWAKFLEL 248
 338 SSSQDCSICMRQSGCLPLIQLLHGNDKDSVLLGNSRSGKEARASAAALHNIHISQPD 397
 249 SGNAQSCATLRRSGCPLLVQMHAFDND-----QEVKCAEQALHNVHSHDE 298
 398 KRGRREIRVLHLEQIOTRAYCETCHEWQEAHEPGM--DQDKNMPAPVEHQICPACVLMK 455
 299 KAGRREAKVRLLDQIVDYCSFKTLTQSGGEIAADSDRHPL-----AALSSLMK 349
 456 LSPDEEHRHAMELGGLOAELQLVDCEMYG-LTNDHYSITLRRYAGMALTNLTFGDVA 514
 350 VSPDEEHRHAMELGGLOAELQLVDCEMYG-LTNDHYSITLRRYAGMALTNLTFGDVA 514
 515 NKATLCMKCMRALVAQLKSESDIQQVITASVLRNLSWRADVNSKKTIREVGSVKALME 574
 410 NKALLCGQKFMALVAQLDSAPDQLQVTASVLRNLSWRADVNSKKTIREVGSVKALME 469
 575 CALEVKESTLKVLSALWNLKSAHCTENKADICAVDGAALAFVGLTLYRSQNTLAIIES 634
 470 AMNRNSENTLKAISALWNLKSAHCTENKAEFCVAVDGAALAFVGLTLYRSQNTLAIIES 529
 635 GGGILRVSSLIATNEDHROILRENNCLOTLLOHLKSHSLTIVSNACGTLWNLKSAHCTEN 694
 530 AGGILRVSSLIATNEDHROILRENNCLOTLLOHLKSHSLTIVSNACGTLWNLKSAHCTEN 589
 695 OEAELWDMGAYMLKNLHSHKMTAMGSAALANLMAANRPAYK-----DANIMSPG-SSL 749
 590 QKFLWDMGAYMLKNLHSHKMTAMGSAALANLMAANRPAYK-----DANIMSPG-SSL 749
 750 PSLHVRKQKALEAELDAQHLSETFDNIDNLSPKASHRSKORHKQSIYGDYVFDNRHDDN 809
 650 PTLKARAKALQELGERHTAETCDNLD-----KIDKERASSSSRRHP----- 677
 810 RSDNFTGNMTVLSPLYNTTVLSSSSSSRGLDSSRSEKORSERLERGICLGNHYHPATEN 869
 678 -----TGG-----KIDKERASSSSRRHP----- 695
 870 PGTSSKRGLOITTAQIAQVMEVSAIHTSQBDRSGSTTELHCVTDERNALRRSSAAH 929
 696 -----APLTR-----SAMLTKSESROSVYSAKSDCAVDH----- 725
 930 THSNYNTKSENRTCSMPYAKLEYKRSNDSNLSSVSSDDGYGKRGQWKPSIESYSED 989
 726 -----LIRSASASDAHRK-----VKPKITDFDLE 749
 990 DESKFCYGVYPADLAHKSANHMDNDGELDTPIYNSLYKYSDEQLNSGRQSPQSNRW 1049
 750 ME-----QTEATEGQPIDYSVKYS----- 769
 1050 ARPKHIIIEIKQEQBSRQSNQSTPYVYTESTDDKHLTKQPHGQOECVSPYRSRGANG 1109
 770 -----ENAKTKSTY----- 778
 1110 SETNRVSGHNGINQVQSLOQEDDYDDKPTNYRSERYSEEQHEBERPTNYSIKYNEE 1169
 779 -----QETDL--DQTPDFSLYAE-----NQTESLDDISGPAGGQ 811
 1170 KRHDQPIDYSLKYATDIP--SSQKQSFSPKSSGQSKTEH-----MSSSENSTSTPSS 1223
 812 KSTITPP-----AETVPEKSEGEIILLDDSVKCYQOTEDTPYVLSNAASVTDLRVAA 864
 1224 NAKQONQLHPS--SAQSRSGOPO---KAATKYSSINQEIOTYCYVEDTPICFSRSSLS 1278
 865 KADREAEVKPEVRVTSKEGAPKLPKLSQCGSGSYTPPEKPINVCEBGTGYSRVDLSLS 924
 1279 SLSSAEDEBIG-CNQTTOBADSANTLOIAEIKETIGTRSAEDPVPSEVPAVSHQHPRTKSSRL 1337

925 SL-----DESKANQAIVGTD-----ADIKPKLEKQEQESQPAEQVLTKPTQANSAL 973
 1338 QGSSLSSESARHKAVERFSSGAKSPKSGAOTPKSPPPHYVQETPLMPFSRCTSVSSLDSE 1397
 974 -----ETPLMFSRRSSMDSLVHDP 992
 1398 SRSTA-----SSVQSEFCGVMVSGIISPSDLPSPGOTMPPSRSKTTPPPQTAQTKREV 1452
 993 DVTVANCDSDSSVSD--FSLASGVISFSEIPDSPTQSMFQS----- 1033
 1453 PKNKAPTAEKRESGPKQAQAAVNAQVRVLPDADTLHFAFATESTPGFCSSLSALSLSLD 1512
 1034 PRRSVAGSGQVDSPPVWIPASLQPLRSVFE--DDLSSFNVEHTPAQFSTATSLNLSI-- 1091
 1513 EPFIQKQVELRIMPVQENDNGNE-----TESEQPKESNENKEKE-AEKTID 1558
 1092 -----VDDEKAPAVWTEDEDELLANCINMGOMKPKTEAVKSTVNVSEVDVAETIR 1144
 1559 S-----EKD-----LLDD--SDDDDI----- 1572
 1145 SYCTEDTPALLSKVPSNTNLVSVSSSTDPKDATAGQAQVMAHQLSDDVSSNASDCGAS 1204
 1573 -EILEECIISAMPTKSSRKAKPAQTAASKLPPVARK-PSQLPVYKLLPSQNRLOPKHV 1630
 1205 GHLQOCTIRDM-----KKPLGEATSDPIAMLRGGNELPGY--LPS----- 1244
 1631 SFTPGDDMPRYVCEGTPINFSTATSLDLTIES-----PPELAAAGGVGRGGAQSGEFK 1686
 1245 -----ADEMK-FLVEDSPCFVSVGSLNLTGSSLVGPAVOLKETE-PSSADQNPENK 1298
 1687 RDTIETGRSTDEAOGGKTSSVTIPELDDNKAEGDILAE----- 1727
 1299 SLANRSRRPPHQDDSLSLSDSDSDTNLLSQAIAAGCNRPKSLGFSNGKRSSL 1358
 1728 -----INSAMPKGSHPKPRVKINDVQQAQASASSAPKQK-----LDG--KK 1769
 1359 SSSOPIAINAATASSLNSAMTVRSQOQESYSSVSDSDNDNOKSLFLCILKGMVKT 1418
 1770 KKP-----TSPVKPIQNTYRT-----RVKNADSKNNAER 1803
 1419 KEPGARAQMOEQPIVSGSSVQSNPSLKQDPSLPVQLPSSQGVKQRHHHHHHHRRER 1478
 1804 VFSNDKSK-KQNLKN--SKDFNDKLPNNEDRVGSGFADSPHYTPIEGTPYCFSRND 1860
 1479 ---ERKDEKLQECINTGISKKIN-AVPKNV--LATSAALEPCH--PMAAT-----T 1523
 1861 SLSSLDLDDDDVLSREKAEALRKAKENKESAKVTSTHTLTSTNOQSANKTQA-TAKQPI- 1918
 1524 SASALSTAPDVE---QKAH--ATSNPQOQSTHPSHILLPNPIDATVTDTRGSPAA 1577
 1919 -NRGQPKPILOKQSTFPOSSKDIIDRGAATDEKLOFNAIEN-----TPVCF----- 1963
 1578 PNOQNGNASQNGLET-ATGSKDLSDSEDSQSFIMETMTVRLDSALNETCISGASEK 1636
 1964 -----SHNSLSLSLSDIDQENNNKENEPIKETEPDPS 1995
 1637 HKDPDLMLKSVERTLMEVFTAELRSSHHSS-----NSHKNNSSNTWNSTCPND 1691
 1996 QGEPSPKQASGVAPKSFHVEDTPVCFNRNSLSLSID---SEDDLOECISSAMPKKK 2052
 1692 VSPFSVSQTA-----PV-----LASLSDDEATEARSHELLEITPTNEQQ 1732
 2053 SRLKGDNEK---HSPRNGGILGEDITLDKIDQRPDSEHGLSPDSENFDMKAIQEGA 2108
 1733 PESLEGETTLVNGHADS--SGSSGGLNQL---GGQVQAGVRLEPQELLENGTSASIMT 1789
 2109 NSIVSSSLHQAACLSRQASDSD-----SILSL-----KSGI-SLGSPP--HUTP 2152
 1790 NSTMIAFEARALAEALLQPAATDDTTMTFSLNSLDLNDIRP?SGMESLNSCYQDSQP 1849
 2153 DQEEKPFSTNKGPRILKPEKSTLETKIESEKIGKGGKVKYSLITGKVRNSETSGQ 2212
 1850 SSLROAMPS-KSPRFAKNFFANLVARALGHLAG-----SAESVNSSCNLDN 1897

530 AGGILNVSHIAVCEPYRQILRHQNCIAILLQOLKBSLTVVNSCGTLWNLSARSAED 589
 695 QBALWDMGAVMLKONLIHSHKHMIAMGSAALNLMANRPAYK-----DANIMSG-SSL 749
 590 QKFLWMDGAVPMLRSLIHSKHAMISEGSSSALKNLINFRPAVONHQHLDPIARSMGLKAL 649
 750 PSLHVRKQKALAEALDAOHLSTFONIDNLSPKASHRSKQRHKQSLYGDYVDFVDTNRHDDN 809
 650 PTLERAKAKALQELGERHTAETCNLD----- 677
 810 RSDNFTNGMNTVLSPLYNTTVLPSSSSRGSLDSRSKXDRSLERBERGIGLGNVHPATEN 869
 678 -----TGG-----KLDKERASSSSRRHP----- 695
 870 PGTSSKRGILQISTTAQIAKWEVEIAHTSOEDRSSGTTTELHCVTDERNALRSSAAH 929
 696 -----APRLTR-----SAMLTKESRDSVYSAKSDCAYH----- 725
 930 THSNTYFTKSENSNRCTCMPYAKLEYKSSNDLSNVSSSDGYGKRGOMKPSIESYSED 989
 726 -----LIRSASASDAHRK-----VKPKITDFDLE 749
 990 DESKPCSYGQYPADLAHKLHSANHMDNDGELDTPIVNSLYKYSDEQLNSGRQSPQNERW 1049
 750 ME-----QDTEATEEQPIDISVKYS----- 769
 1050 ARPKHIIIEIKQEQORQSRNQSTTVPVYTESTDDKHLKFPQHQEQCVSPYRSGANG 1109
 770 -----ENATKSTV----- 778
 1110 SETNRVGSNGINQVNSQSLQEDDEDKPTNYSERYSEEBEERPTNYSIKYNEE 1169
 779 -----QETDL-----DQPTDFSLRYAE-----NOIESDLOISGAPAGQ 811
 1170 KRVHDOPIDYSLKYATDIP--SSQKQSFSSKSSGQSSKTEH-----MSSSENTSTPS 1223
 812 KSTITPP-----AETVPEKSEGELEILLDSDVKCYCTEDTPYVISNAASVTLDRVA 864
 1224 NAKRONQLHPS--SAQSRSGOPO--KAATCKVSSINQETIOTYCYVEDTPICFSRSSLS 1278
 865 KADAEABVKEPREVTSKEGAPKLPKLSQCGSGSYTPEKPINYCEGTPGVFSRYDLS 924
 1279 SLSSAEDEIG-CNOTQEQASANTLQABIKETGRSAEDPVSEVPVAVSQHPRTKSSRL 1337
 925 SL-----DESKANQAIUGTD-----ADIKPLEKQEESOPAEQVLTKPTQANSAL 973
 1338 QGSSLSSESARHKAVFSSGAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVSSLDSP 1397
 974 -----ETPLMFSRRSSMDSLVHDP 992
 1398 SRSTA-----SSVQSEPCSGMVGILISPSDLPDSRGOTMPPSRKTPPPPPQATQKREV 1452
 993 DVDVANDCKSSVVSD-FSLRAGSVISPEISDPTQSMPOG----- 1033
 1453 PKNKAPTAEKRESGPKQAANAAVQRVQLPDADTLHFAESTPDGFCSSLSALS- 1511
 1034 PRNSVAGSGQVNDSPWPVIPSILQPLRSVFE-DLSSFNVEHTPAQFSTATSLNLSIV 1092
 1512 -DEFFIOKVELRIMPVQENDNGNE-----TESQPKESNENQKE-AEXT 1556
 1093 DDE-----KAPASVAEEDNEDELLLANCINMGORMKPTAEVKSIVVNSVDVAET 1143
 1557 IDS-----EKD-----LLDD--SDDDDI-- 1572
 1144 IRSYCTEDTPALLSKVPNTNLSVISMSTDPKDATAGQAMVAHQSLSDVSSNASDCGG 1203
 1573 ---BILBEICISAMPTKSSRKAQKAPQATASKLPPPVARK-PSOLVYKLLPSONRLOPQK 1628
 1204 ASGHLLQOCIRDMG-----KKPLGEATSDPTAMLRGNGELPGY--LPS----- 1245
 1629 HVSTPGDMPRVYCVSGTPINRSTATSLDLTIES-----PPNELAAGEGVGGAQSGEF 1684
 1246 -----ADEMNK-FLVEDSPCNFVSGLNLTVGSSLVGPAVLKETE--PSSADQNPEM 1297

1685 EKRDITPTEGRSTDEAQGGKTSSVTIPELDDNKAEEGDILAE----- 1727
 1298 KAKPGKEQVRRPHWQDDSLSLSDSEDDTNLLQAAGCNRPKSNLGFSSNGKRSS 1357
 1728 -----INSAMPKQKSHKPPRVKIMDOVOQASASSAPKNQ-----LDG-- 1767
 1358 SLSSSQPIAINAATSASSLNSAMTVRKSOQOESYSSVSDSDNDNQSLSLFCILKGMV 1417
 1768 KKKKP-----TSPVKPIQNTET-----RVKNADSKNLANA 1801
 1418 KKEFGARAAQMOEQPIVSGSSVQSNPSLKQPDLPVQLPSSGOVKQRHHHHHHHRR 1477
 1802 ERFVSDNOKK-KONLKQN--SKDFNDKLPNEDRVGSPAFSPHYPHYPTEICTFCFSR 1858
 1478 ER---ERKDEKLQECINTGISKKIN-AVEKNV--LATSAAALEPCH--PMAAT----- 1523
 1859 NDSLSLDFDDDDVDLSREKAEKAKENKESAKVTSTELTNSQOSANKTOAIKQPI 1918
 1524 -TSASALSTAAPV-----EOKAHATSNPQ---QOSSTHPSHILPNPI 1563
 1919 -----NRQPKPILQKQSTFPPOSSKDIPDRGAATDEKLOFAIEN----- 1958
 1564 DAIAVTDTVRSAPAPNQGNGNASQNGLET-ATGSKDLDESDRSDSNQSFIMETWRL 1622
 1959 -----TPYCF-----SHNSLSLSLSDIDQENN 1980
 1623 DSALNETCISGASEKHKDDPDLMLKSVERTWFTVSAEQLRSSSHHSS-----NSHKN 1677
 1981 NKENEPIKETEPDPSQGEPSKQASGYAPKS FHVEDTPVCFSRNSSLSSLSID---SEDD 2037
 1678 NSSNNTWNESTCPNDVSPSVQTA-----PV-----LASLSLDEDATEAR 1718
 2038 LIQECISAMPKKKPSRLKGDNEK---HSPRNMGGILGEDLTLDLKDQRPDSEHGLS 2093
 1719 SLHELIEITPNEQSPLEGETDILVNHADSVSGSSGGLNQL---GGQVQVAGVLE 1775
 2094 PDSENFOWKATQEGANSIVSLHQAACLSROASDSD-----SILSL-----KS 2140
 1776 PQLLFGTSSASITNSTWIAFEARALAEALQAPATDDTTTMTFSLNLDLNRIPPS 1835
 2141 GI-SLGGFF--HLTPDQEEKFTSNKGPRLKPGKSTLETKIESKGIKGGKVKYS 2197
 1836 GMEISLNCYQDHSQFSSLRQAMPS-KSPRFARKVFPANLVARRALHLAG----- 1884
 2198 LITGVRSNSISGOMQOP-----LOANMPSISRGRTMIHIPGVNRSSSSTSPVSKKGP 2252
 1885 -SAESVNSSCNLLDNIKPPLMDELDSMISVDSIQSEVADGEQDCSMATTISVSNYETA 1943
 2253 L-----KTPASKSPSEGTATTSPRGAKPSVKSEL----- 2282
 1944 ACDDQMTVLQSCFDEDEDATMNDYSSAESTPKHGSTFSPNRRSLTPKDRRLTKDRFKT 2003
 2283 -----SPVARQTSOI-----GGSKAPSRGSRDSTPSRPAQOPLSRPI--- 2321
 2004 YTIATSCMEAPEANETLQIEIVAAVPAVTPSPRANGRRGSAERYKTOLIECPALIQ 2063
 2322 -----QSP---GRNISI3GR 2333
 2064 PQDDCPSEQLSSITRAMMQOFTFITDINIGHSQETCESTDHPEDAGESPECQDNETESC 2123
 2334 NG-----ISPENKLSOL-----PRTSSPSTASTKSSGSGKVSYTPGRQWSQONLTKQT 2382
 2124 DQGEQDQPPPPSIVDLRTSVKPTTLEPAVAKLVGRKKPAVSVYSWMSQON----- 2178
 2383 GLSKNASSIPRSESASKGLNQWANGNGANKVELSRMSSTKS-----SGSESDSRSEPV- 2436
 2179 -----SNNAPSKKKTLSPTIAKESLVPGGSGVRLPAKAKPTP 2216
 2437 -----LVQSTFIKEAPSTLRLKLEESASFESLSPSSRPASPTESQATVLSLPSL 2488
 2217 PPEPAPARLERQGTFKD-----EPTNSNVQVPVVE-TK 2249

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sigen-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Yu X., Bienz M.;
RT "A new *Drosophila* APC homolog concentrated in apical adhesion zones of
RT epithelial cells,"
RL Nat. Cell Biol. 0:0-0 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99147086; PubMed=10021369;
RA van Es J.H., Kirkpatrick C., van de Wetering M., Molenaar M.,
RA Miles A., Kuipers J., Destree O., Peifer M., Clevers H.;
RT "Identification of APC2, a homologue of the adenomatous polyposis coli
RT tumour suppressor,"
RL Curr. Biol. 9:103-108 (1999).
DR EMBL; AF003746; AAF56249.1; -
DR EMBL; AF113913; AAD40227.1; -
DR EMBL; AF091430; AAD20985.2; -
DR FlyBase; FBgn0026598; Apc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 6.
DR PROSITE; PS01176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116702 MW; 20C8F5F6121888F8 CRC64;

Query Match 8.4%; Score 1222.5; DB 5; Length 1067;
Best local similarity 26.3%; Pred. No. 8.6e-52;
Matches 429; Conservative 178; Mismatches 379; Indels 647; Gaps 50;

QY 329 DMSRTLLAMSSQSCISMROSGCLPLLIQLLHGNKDSVLLGNSRGSKARARASAAH 388
DB 12 ELTRNFELSNPECTALRSSDCIQLLVQILHANDE-----GLSTAKKYASQALH 62
QY 389 NIHSQPDKGRREIRVHLLEQIRAYCETC-NEWQEAHP-GWDOQKNMPAPVEHQI 446
DB 63 NIVHNPEKQRQREVMKRLLDQILDYCNFLHTQLQSGGEAIAADDEDRHPL----- 114
QY 447 CPACVLMKLGDFDEHRHAMELGLOAIABLLQVDCBMVG-LTNDHYSITLRYAGVAL 505
DB 115 -AAMKLLMKASFDEHRTQNCGLNKAIFNLVHLDAVHGPAAGRCQCALRYGVAL 173
QY 506 TNLTFGD--VANKATLCMSKGMALVAQLKSEDELOQVIASVRLNLSWRADVNSKTL 563
DB 174 TNLTFGDENVNKSQCGQFMVIAQLNAPDELQVLQVLNLSWRADKMKMTIF 233
QY 564 REVGSVKALMECALEVKKESTLKVLSALWLSAHTENKADICAVDGAFLVGLTYR 623
DB 234 NELGTVTSLAARQNKENTLKALLSALWLSAHCSTNKAFCVADGALAFVGLMSYE 293
QY 624 SQTNTLAIIESGGGILRNVSIIATNEDHRIILRENCLQTLLOHLKSHSITIVSNACGT 683
DB 294 GPSKTLKIENAGGILRNVSIIAVCEPYRIILRYNCLAILLQQLKSESITIVSNSCGT 353

QY 684 LWNLSARNPKDQBALWDMGAVSMLKNLIHSHKMIAMGSAALRNLMANRPAYKDNAM 743
DB 354 LWNLSARCPEDQYLLDHNPAIPILRALISSKNMIAEGSASALKNLVNFATBELMPN-- 411
QY 744 SPSSPSLHVRKQKALEAEADACHLSETPDNIDNLSPKASHRSKORHKOSLYGVYFDT 803
DB 412 GDGGSLLP----- 418
QY 804 NRHDDNRSDNFNTGNMTVLSPLYLNTTVLPSSSSRGLSDSRSEKORSLEEREGIGLGN 863
DB 419 -----LDKEAGHG----- 426
QY 864 HPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTSQEDRSSGSTTELHCVTDERNALR 923
DB 427 -----GTLPRR----- 432
QY 924 RSSAAHTHSNTYNTFKSERNRSCMPYAKLEYKRSNDLSNLSVSSDGYGKGQMKPSI 983
DB 433 -----FSSLRLSSNPTGSLKKVRFS-----TVSTTGFNLNKKSGRESI 470
QY 984 ESYSEDESKFCSYGQYPADLAHKIHSANHMDNDGELDTPINYSLKYSDEQLNSGRQSP 1043
DB 471 YSGKSD-----STKYST----- 482
QY 1044 SQNERWAPKHIIIEIKOSEQRQSNQSTTPYVYTESTDDKHLKFPQHGQOECVSPYR 1103
DB 483 -----K 483
QY 1104 SRGANGSETNRVGSNHEINQVNSQLCQEDDYDDDKPTNYSERYSEEQHEEBERTNYS 1163
DB 484 SEG-----KNPFEIVTPT-----EEQIDYDS 505
QY 1164 IKNEKRRH-----VQPIDYSLKYATDIPSSQKQSFSSKSSGSSGSSKTEHMSSES 1215
DB 506 MKYMEHKPNSSKTFEIDLQPTDFSARY-----KERRSAQTPAKELASSETN 551
QY 1216 ENTSTPSSNAKRONLHPSSAQSR-----SQCPKAATCKVSSINQETIQYCYVE 1265
DB 552 EIRS-----KELQTKSSATLNSPGLVAVSAKOKIAT-ETETETAEPINYCEE 603
QY 1266 DTPICFSCRSLSLSLSAEDIGCNOTQOASANTLQIAEIKETIGTRSAEDPVSEVPA 1325
DB 604 GTPGFSFRPDSLNSL-----TEKPE 623
QY 1326 VSQHPPTKSSRLQSGSLSSSARHKAVEFSSGAKSPSKGAQTPKSPPEHYVQETPLMFS 1385
DB 624 KCMPPKTPKT-----AVLEPHVDGNTPOIDS--ALETFMFS 660
QY 1386 RCTSVSLSDFSERSIA-----SSVQSEPCSGMVSGIISPSDLPDSPGQTMP--PSRSKTP 1439
DB 661 RRSMDSLVG-DDETVAECEDNGSVISE-YRMSQSGVISELPDPTQSMQSPRRDR-- 716
QY 1440 PPPPQTAQTKREVPRKNAPTAEKRESGPKQAAVAAVQVQLPDDATLLHFATESPTDG 1499
DB 717 ---KVSTQNNLDTPEQKPTVFE-----DKLNRPHVEHTPAA 750
QY 1500 FSCSSLSLSALSDPEPFIQXDVLRIMPVQENDNGNETSEQPKESNCE-KEAETID 1558
DB 751 FSCATLSLSLSM-----MDSNANAIQGRGNDINGMDAPRSTCTED 793
QY 1559 SEKDLDDDDDDIEILEECIIISAMPTKSSRKAQPAQASKLPPVPA-----RKPSQ 1611
DB 794 TTAVALSKAPNSDLSIL-----SIPNDLN-----ANEAQVPVAPRADVTGMDTMAPAE 841
QY 1612 LPVYKLLPQNRLQPKHYVFTPGDDMPVYCVGEGTPIFNSTATSLSDLTIESPPNELAA 1671
DB 842 DAISKVRCGGNALP-----SYLPVSDMSKYYVESPTCTFSVIGLSHUTV----- 887
QY 1672 GEGVRGAQSGEFKRDITPTIGRSTDEAQGGKTSVTTPPELDNKK-ABEGDILAEICNS 1730
DB 888 -----GSAKAGPVLK--LPM--RTAEEAQA-----PKLPPRESAVCGD-----AEP 924
QY 1731 AMPKXGSHKPFVRKKINDQ-----VQOASASSAPNKNQDLGKKKFTSPVKPIQONTYR 1786


```

Db 925 RLPPKSDLSLSSMSDDDDCNLLSQAAGSC-----RQPSGA--- 964
Qy 1787 TRVRKNADSKNNLNAERFSDNKKDKQKLNKNSKDFNDKLFNEDRVGSAFDPSPHY 1846
Db 965 -----STSSSLANASTLCLRENGQSKQ-----VEHGDK-FN----- 996
Qy 1847 TPIEGPYCFSRNDSLSLDFDDDDVLSREK-----AELRAKENKSEA----- 1892
Db 997 -----YSSDLSL-----DDDDDAKSKLFEQCILSGHKSNDALSEGEPPGQR 1041
Qy 1893 -KVTSHTELTNQ 1904
Db 1042 QEISARDRFVSNQ 1054

RESULT 13
Q961B0 PRELIMINARY; PRT; 1067 AA.
AC Q961B0;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE LB24920P
GN APC2 OR CG6193.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Eohydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celnikier S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051719; AAK93143.1;
DR FlyBase; FBGN026598; Apc2.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR SMART; SM00185; ARM; 7.
DR PROSITE; PSS0176; ARM_REPEAT; 1.
SQ SEQUENCE 1067 AA; 116678 MW; 03C4119AEF19D198 CRC64;

Query Match 8.4%; Score 1221.5; DB 5; Length 1067;
Best Local Similarity 26.3%; Pred. No. 9.7e-52;
Matches 429; Conservative 178; Mismatches 379; Indels 647; Gaps 50;

Qy 329 DMSRTLLAMSSQDSCTSMQSCGLPLLIQLLHGNKDKSVLLGNRGSKEARASAAHL 388
Db 12 ELTRNLELGRNPBTCTALSSDCIQLLQILHANDE-----GLSTAKYASQALH 62
Qy 389 NIIHQSPDDKRGREIRVLHLEQIRAYCTC-WEQEAHEP-GMDQKNMPEAPVEHQI 446
Db 63 NIVHNPEERQREVKMLDLQILDYCNFLHTQ-JSGGEALADDEDRPL----- 114
Qy 447 CPAYCVLMKLUSFDEHHRAMNEGLGQIAELQLQVDCMYG-LTNDHYSITLRRYAGMAL 505
Db 115 -AAMKLLMKASFDEHRTQTCELGALKAIPLNVLHDAVHGAAGMEQCNCALRSYGLMAL 173
Qy 506 TNLTFGD--YANKATLCSMGCEALVAOLKSESDIQQVIAVLRLNLSWRADVNSKTL 563
Db 174 TNLTFGDENVHNSYLCQQQOFMEVTAQNTAPDELLQVLAVLRLNLSWRADNKHMTIF 233
Qy 564 REVGSVKALMECALVKKSTLKSLSALWNLNLSAHTENKADICAVDGAFLAVLGLTYR 623
Db 234 NELGTVTSLARAAQNKNTNLTAKLSALWNLNLSAHTENKAEFCVADGALAVLGLMSYE 293
Qy 624 SQNTLAIISGGGILRNVSLSIATNEDHQILRENNCLQTLQLHLKSHSLTIVSNACGT 683

```

```

Db 294 GPSKTLKIIENAGGILRNVSLSIAVCEPYRQILRYNCLAILLQOLKSESLTVNSCGT 353
Qy 684 LWNLSARNPKDQBALWDMGAVSMKLNLIHSHKHMVAMGSAALRNLMANRPARYKADANIM 743
Db 354 LWNLSARNPKDQBALWDMGAVSMKLNLIHSHKHMVAMGSAALRNLMANRPARYKADANIM 743
Qy 744 SPGSSLSPLSHVRKQKALEAELDAQHLSETPDNDNLSPKASHRSKORHKOSLYGVYEDT 803
Db 412 GDGGLP----- 418
Qy 804 NRHDDNRSDNFTGNMTVLPYLNTTVLPSSSSSRGSLDSSSRSEKORSERERIGLGN 863
Db 419 -----LDKEAGHG----- 426
Qy 864 HPATENPGTSSKRLQISTTAAQIAKYMEVSAIHTSQEDRSSGSTTELHCVTDERNALR 923
Db 427 -----GTLPRR----- 432
Qy 924 RSSAAHTSHNTYNTKSENRTCSMPYAKLEYKRSNDSLSNVSSSDGKGKGMKPSI 983
Db 433 -----FSSLRSSNFTGSLKKVRPS-----TVSTTGLNRRKCESRESI 470
Qy 984 ESYSEDDDESKFCSYGYQYPADLAHKAHSAHMDNDGELDTPIYNSLYKYSDEQLNSCRQSP 1043
Db 471 YSGKSD-----STKYST----- 482
Qy 1044 SQNERWAPKHIIEDEIKQSEQRQSRNQSTYPVYTESTDDKHLKFPQHPGQECVSPYR 1103
Db 483 -----K----- 483
Qy 1104 SRGANGSETNRVGNHGINGNVQSQCQEDDYEDDKPTNYSEYSEBEQHEEBERPTNYS 1163
Db 484 SEGA-----KNPFEIVT-----BEQPIDYS 505
Qy 1164 IKYNEEKRH-----VDQPIDYSLKYATDIFSSQKQSFQSKSSGQSKQSKTEHMSSS 1215
Db 506 MKYMEHPNSKTFEIDLDTQDPSARY-----KERRSAQTAQPELASETN 551
Qy 1216 ENTSTPSNAKRONLHPSSNAQR-----SGOPKAACTKVSSINQETIQTVCVE 1265
Db 552 EIRS-----KELQTKSSSATELRNLSPLGVAASAQKQIAT-ETETETAEAPINCEE 603
Qy 1266 DTPICFCGRCSLSLSLSEAEDIEGNCQTQTEADSNLTQIAIEIKICIGTSAEDPVSVP 1325
Db 604 GTPGSFSEFDSLSL-----TEKPE 623
Qy 1326 VSQHPRTKSKSLQSSLSLSSSSARHKAVEFSSGAKSPSKGAQTEPKSPPEHVQETPLMFS 1385
Db 624 KMPPTKPTKT-----AVLPVHVDGNTFQNIIDS-ALETPLMFS 660
Qy 1386 RCTSVSSLDSPESRIA-----SSVQSPPCSGMWGII-SPDLDPSPQGTWP--PSRSKTP 1439
Db 661 RRSMDLSLVG-DDETVACEDNGSVISB-YSRMQSGVISPSELDPSPQSMFQSPRRDR-- 716
Qy 1440 PPPPQATKKEVPKNKAPTAKRESGPKAANVAQVRVQVLDPADTLHFATSTPDG 1499
Db 717 ---KVSTQNNLDTPEQKPTVFE-----DKLARFHVHEHPAA 750
Qy 1500 FSCSSLSALSLSLDEPFQIKDVELRIMPVQENDNGNETESQPKESNENOB-KEAEKTI 1558
Db 751 FSCATLSLSLSM-----MDDSNANIRGORGNDINGNGDAPSYCTED 793
Qy 1559 SEKDLDDSDDDDIIEECIISAMPTKSKSKAKKPAQTASKLPPVPA-----RPSQ 1611
Db 794 TTVAVLSKAPNSDL-SIL-----SIPDLN-----ANEQFVAPAPADVTDGMDTRPAE 841
Qy 1612 LPVYKLLSQNLQPKQHVSTPGDMPRVYCVGTPINFSTATSLDLSLTTESPPNELAA 1671
Db 842 DAISKORCGGNALP-----SYLPVSDENSKYVEDSPCTFSVISGLSHLTV----- 887
Qy 1672 GEGVGGAGQSSEFEKRDITPTTEGRSTDEAQGGKTSSVTIPELDNDK-ABEGDILAECS 1730

```

Db 888 -----GSAKAGPVILK---LPM--RTAEAAQA-----PKLPPRRSAVOGD-----AEP 924
QY 1731 AMPKGSHKPPFRVKKIMDO-----VOASASSAPNKNQDGGKKKPTGPVKIPQNTBYR 1786
Db 925 RLFPKXSDLSLSLSDSDCCNLLSQIAAGSC-----RPQPSGA--- 964
QY 1787 TRVRKNADSKNINLAERVPFSDKKNLKNNSXDENDKLPNNEDVRGSPAFDPSPHY 1846
Db 965 -----STSSSLANASTILCRENGQKKQ-----VEHGDK-PN----- 996
QY 1847 TPIETPYCFNRNDSLSLDFDDVDLSREK-----AELRKAKENKSEEA----- 1892
Db 997 -----YSSDLSL-----DDDDDDARSLSLFEQCILSGMKNSDALESEGBPPQQR 1041
QY 1893 -KVTSHTELTNSQ 1904
Db 1042 QEISARDRFVSNQ 1054
RESULT 14
Q9P119
ID O9P119 PRELIMINARY; PRT; 208 AA.
AC O9P119,
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adenomatosis polyposis coli tumor suppressor (fragment).
GN APC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435068; PubMed=10982189;
RA Su L.K., Steinbach G., Sawyer J.C., Hindi M., Ward P.A., Lynch P.M.;
RT "Genomic rearrangements of the APC tumor-suppressor gene in familial
RT adenomatous polyposis";
RL Hum. Genet. 106:101-107(2000).
DR EMBL; AF127506; AAF34355.1; -.
DR EMBL; AF127034; AAF34355.1; JOINED.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 2.
FT NON_TER 1
FT NON_TER 208
SQ SEQUENCE 208 AA; 22726 MW; 51076F257B61C528 CRC64;
Query Match 7.1%; Score 1041; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 8.1e-44;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 471 GLQAIAEQLQVDCMYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCSMKGCMRALV 530
Db 1 GLQAIAEQLQVDCMYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCSMKGCMRALV 60
QY 531 AOLKSESEDLQOVIASVLRNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVL 590
Db 61 AOLKSESEDLQOVIASVLRNLSWRADVNSKTLREVGSKALMECALEVKKESTLKSVL 120
QY 591 ALWNLSAHCTENKADICAVDGALAFVGLTLYRSQTNLTALIESGGGILRNVSLLIATNE 650
Db 121 ALWNLSAHCTENKADICAVDGALAFVGLTLYRSQTNLTALIESGGGILRNVSLLIATNE 180
QY 651 DHRQILRENNCLOTLLQHLKSHSLTIVS 678
Db 181 DHRQILRENNCLOTLLQHLKSHSLTIVS 208
RESULT 15
Q7Z2Q8
ID Q7Z2Q8 PRELIMINARY; PRT; 159 AA.
AC Q7Z2Q8;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056268; AAHS6268.1; -.
KW Hypothetical protein.
FT NON_TER 159
FT NON_TER 159
SQ SEQUENCE 159 AA; 17843 MW; E0C0CC055A22C91B CRC64;
Query Match 5.2%; Score 756; DB 4; Length 159;
Best Local Similarity 98.7%; Pred. No. 5.9e-30;
Matches 153; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAASYDQLLKQVEALKMENSRLRQELDNSNHLTKLETEASNKKEVLKQLOGSIDEAM 60
Db 1 MAAASYDQLLKQVEALKMENSRLRQELDNSNHLTKLETEASNKKEVLKQLOGSIDEAM 60
QY 61 ASSGQIDLLERLKEMLNDSNPFPGVKLRKMSLRYSYSGSREGSVSSRGCSPPVPMGSFPR 120
Db 61 ASSGQIDLLERLKEMLNDSNPFPGVKLRKMSLRYSYSGSREGSVSSRGCSPPVPMGSFPR 120
QY 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEK 155
Db 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEK 155
Search completed: August 25, 2004, 17:23:07
Job time : 214.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 16:58:01 ; Search time 180.5 Seconds

(without alignments)
4450.319 Million cell updates/sec

Title: US-09-442-489f-7

Perfect score: 14566

Sequence: 1 MAASVDQLLQVEALKMEN.....ESSGTQPKRHSGLVLTSTV 2843

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14566	100.0	2843	2 AAW76140	AAW76140 Human APC
2	14566	100.0	2843	2 AAW76144	AAW76144 Human APC
3	14566	100.0	2843	3 AAB23011	AAB23011 Human APC
4	14566	100.0	2843	5 ABG71105	ABG71105 Human ade
5	14566	100.0	2973	2 AAW76821	AAW76821 Human APC
6	14566	100.0	2973	4 AAW72782	AAW72782 Transcript
7	14559	100.0	2973	3 AAY70304	AAY70304 Protein u
8	14548.5	99.9	2842	2 AAR63508	AAR63508 Adenomat
9	14548.5	99.9	2842	5 AAG90968	AAG90968 Human APC
10	14546	99.9	2843	2 AAW11922	AAW11922 Adenomat
11	14539	99.8	2843	7 ADE65846	ADE65846 Human ade
12	14533	99.8	2843	2 AAR26052	AAR26052 APC gene
13	14533	99.8	2843	2 AAW35392	AAW35392 Human ade
14	14533	99.8	2843	2 AAW38370	AAW38370 Human ade
15	14533	99.8	2843	5 ABG90964	ABG90964 Human ade
16	14509	99.6	2860	2 AAR63507	AAR63507 Adenomat
17	14479	99.4	2843	2 AAR58634	AAR58634 Adenomat
18	13999.5	96.1	2742	3 AAB23012	AAB23012 Human APC
19	13198.5	90.6	2842	7 ADE56175	ADE56175 Rat Prote
20	4738	32.5	912	5 ABG71106	ABG71106 Human ade
21	3892	26.7	767	5 ABG71107	ABG71107 Human ade
22	3512.5	24.1	2274	4 AAB50674	AAB50674 Mouse APC
23	3488.5	23.9	902	4 AAG09935	AAG09935 Novel hum
24	3415.5	23.4	2303	6 ABR58648	ABR58648 Human can
25	2169.5	14.9	799	3 AAY92061	AAY92061 Human APC

26	1773.5	12.2	1674	3 AAY92060	AAY92060 Murine AP
27	1765	12.1	2417	4 ABB58126	ABB58126 Drosophil
28	1490	10.2	332	2 AAR88353	AAR88353 Mutant A-
29	1222.5	8.4	1067	4 ABB62156	ABB62156 Drosophil
30	562.5	3.9	2344	4 AAU37120	AAU37120 Staphyloc
31	537	3.7	2768	4 ABB68397	ABB68397 Drosophil
32	527.5	3.6	3257	4 ABB67502	ABB67502 Drosophil
33	517.5	3.6	2586	4 ABB66878	ABB66878 Drosophil
34	517	3.5	2271	6 ABU16000	ABU16000 Protein e
35	516	3.5	2271	6 ABM72734	ABM72734 Staphyloc
36	516	3.5	2283	6 ABP56876	ABP56876 Staphyloc
37	515	3.5	2261	6 ABJ18914	ABJ18914 Pathogen
38	511.5	3.5	1186	4 AAB50654	AAB50654 C. elegan
39	507.5	3.5	178	2 AAW33894	AAW33894 Flea sali
40	507.5	3.5	178	2 AAW82368	AAW82368 Flea sali
41	502.5	3.4	2137	5 ABP39618	ABP39618 Staphyloc
42	492	3.4	3111	4 ABB60327	ABB60327 Drosophil
43	478.5	3.3	2951	4 ABB60291	ABB60291 Drosophil
44	474	3.3	5533	4 ABB65772	ABB65772 Drosophil
45	474	3.3	5560	4 ABB71160	ABB71160 Drosophil

ALIGNMENTS

RESULT 1

AAW76140
ID AAW76140 standard; protein; 2843 AA.

XX AC AAW76140;

XX DT 25-MAR-2003 (revised)

DT 23-NOV-1998 (first entry)

XX DE Human APC protein #1.

XX KW Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;

KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
Gardner's Syndrome; GS; predisposition.

XX OS Homo sapiens.

XX PN US5783666-A.

XX PD 21-JUL-1998.

XX PF 25-MAY-1995; 95US-00452655.

XX PR 16-JAN-1991; 91GB-00000962.

PR 16-JAN-1991; 91GB-00000963.

PR 16-JAN-1991; 91GB-00000974.

PR 16-JAN-1991; 91GB-00000975.

PR 08-AUG-1991; 91US-00741940.

PR 12-AUG-1994; 94US-00289548.

XX PA (CANC-) CANCER INST.

PA (UYUO) UNIV JOHNS HOPKINS.

PA (UTAH) UNIV UTAH.

XX PA (ZENE) ZENECA PHARM.

PI Kinzler K, Joslyn G, Markham AP, Carlson M, White RL;

PI Thliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;

PI Vogelstein B, Hedge PJ;

XX WPI; 1998-427100/36.

DR N-PSDB; AAV56447.

XX Adenomatous polyposis coli protein - useful in the treatment of cancers

PT associated with mutation(s) on human chromosome 5q21.

XX Disclosure; Col 41-54; 102pp; English.

XX This sequence represents a human familial adenomatous polyposis coli

CC

CC (APC) protein from clone pp2.5. The gene for the protein is present on
 CC human chromosome 5q21 and is also referred to as adenomatous polyposis
 CC coli gene. It is a tumour suppressor gene, and mutations in this gene
 CC have been associated with tumorigenesis in retinoblastoma and colorectal
 CC tumours, and especially familial adenomatous polyposis (FAP) and
 CC Gardner's Syndrome (GS). The protein can be used in therapy to replace
 CC lack of native functional protein and the nucleic acids can be used for
 CC gene therapy. The nucleic acids that encode them can also be used as
 CC probes and primers in detection of the cancers and predisposition to it.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC
 XX
 SQ

Sequence 2843 AA;

Query Match 100.0%; Score 14566; DB 2; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAASVDLLKQVEALKWENLNRLQELNNSHLTKLETEASNMKEVLKQLOGSIEDEAM	60
DB	1	MAAASVDLLKQVEALKWENLNRLQELNNSHLTKLETEASNMKEVLKQLOGSIEDEAM	60
QY	61	ASSQIDLLRLKELNLDSSNFFGVKLRKMSLRSGSREGSVSSRSRSGCSPVPMGSPFR	120
DB	61	ASSQIDLLRLKELNLDSSNFFGVKLRKMSLRSGSREGSVSSRSRSGCSPVPMGSPFR	120
QY	121	RGFVNGSRESTGYLELEKERSILLADLKEEKEKDWYQAOLNLTKRIDSPLTENFSL	180
DB	121	RGFVNGSRESTGYLELEKERSILLADLKEEKEKDWYQAOLNLTKRIDSPLTENFSL	180
QY	181	QDWTNRQLYEARQIRVAMEEOLGTQCDMEKQAQRRIARIQIIEKDILRIQLQSQAT	240
DB	181	QDWTNRQLYEARQIRVAMEEOLGTQCDMEKQAQRRIARIQIIEKDILRIQLQSQAT	240
QY	241	EAESSQNKHETGSHDAERONEGGVGEINMATSGNGQSGTTRMDHETASVLSSSSTHSA	300
DB	241	EAESSQNKHETGSHDAERONEGGVGEINMATSGNGQSGTTRMDHETASVLSSSSTHSA	300
QY	301	PRRLTSLGTGKVMYVSLLSMLGTHDKDMSRTLLAMSSQDSCIENRSGGCLPLLIQLL	360
DB	301	PRRLTSLGTGKVMYVSLLSMLGTHDKDMSRTLLAMSSQDSCIENRSGGCLPLLIQLL	360
QY	361	HGNDKDSVLLNGSRGSKAEARASAAHLNIHSDPDKRGRRIRVLHLEQIRAYCETC	420
DB	361	HGNDKDSVLLNGSRGSKAEARASAAHLNIHSDPDKRGRRIRVLHLEQIRAYCETC	420
QY	421	WENQEAHEPGMDQKNPMPAPVHQICPAVCVLMKLSFDEEHRHAMELGGLOAIAELLQ	480
DB	421	WENQEAHEPGMDQKNPMPAPVHQICPAVCVLMKLSFDEEHRHAMELGGLOAIAELLQ	480
QY	481	VDCEMYGLTNDHYSITLRRYAGKALTNLTFGDVANKATLCSMGKCMRVALVAQLKSEEDL	540
DB	481	VDCEMYGLTNDHYSITLRRYAGKALTNLTFGDVANKATLCSMGKCMRVALVAQLKSEEDL	540
QY	541	QOVIASVLRNLWRADVNSKTLREVGSKVAMECALEVKKESTLKSVALNLSAHTC	600
DB	541	QOVIASVLRNLWRADVNSKTLREVGSKVAMECALEVKKESTLKSVALNLSAHTC	600
QY	601	ENKADICAVDGAALFLVGTITYSQNTLAIIESGGGILRNVSLLATNEDHRQLRENN	660
DB	601	ENKADICAVDGAALFLVGTITYSQNTLAIIESGGGILRNVSLLATNEDHRQLRENN	660
QY	661	CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSLKLIHSHKHVMIAM	720
DB	661	CLQTLLOHLKSHSLTIVSNACGTLNLSARNPKDQALWDMGAVSLKLIHSHKHVMIAM	720
QY	721	GSAALRNLMANRPAPKADANIMSPGSSLSLVRKQKALEAELDAQHLSSETFDNIDNLS	780
DB	721	GSAALRNLMANRPAPKADANIMSPGSSLSLVRKQKALEAELDAQHLSSETFDNIDNLS	780
QY	781	PKASHRSKQRHKSGLYGDYVFDNRHDDNRSDFNTGNMTVLSFYNTTLPSSSSSRGS	840
DB	781	PKASHRSKQRHKSGLYGDYVFDNRHDDNRSDFNTGNMTVLSFYNTTLPSSSSSRGS	840

QY	841	LDSSRSRSEKRSLESERGIGLGNVHPATENPFTSSKRGLOISTTAAQIAKVNBEVSAIHTS	900
DB	841	LDSSRSRSEKRSLESERGIGLGNVHPATENPFTSSKRGLOISTTAAQIAKVNBEVSAIHTS	900
QY	901	QEDRSSGSTTTLCHVCTDERNALRESSAAHNTSNYNTFTKSENSNRCTCMPYAKLEYKRSS	960
DB	901	QEDRSSGSTTTLCHVCTDERNALRESSAAHNTSNYNTFTKSENSNRCTCMPYAKLEYKRSS	960
QY	961	NDSLNSVSSSDGYGKRGCMKPSIESYSEDESKFCSYGYPADLAHAKIHSAHMDNDGE	1020
DB	961	NDSLNSVSSSDGYGKRGCMKPSIESYSEDESKFCSYGYPADLAHAKIHSAHMDNDGE	1020
QY	1021	LDTINYSKYSDEQLNSGRQSPQNERWARPKHIIIDEIKQSEQRQSRNQSTTYPVYTE	1080
DB	1021	LDTINYSKYSDEQLNSGRQSPQNERWARPKHIIIDEIKQSEQRQSRNQSTTYPVYTE	1080
QY	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGNHGINQNVNSQSLCOEDDYEDDKP	1140
DB	1081	STDDKHLKFPHFQOQECVSPYRSRGANGSETNRVGNHGINQNVNSQSLCOEDDYEDDKP	1140
QY	1141	TNYSERYSSEHQHEBERPTNYSIKYNEEKHVDQPTDYSLKATDIPSSQKQSFSEKS	1200
DB	1141	TNYSERYSSEHQHEBERPTNYSIKYNEEKHVDQPTDYSLKATDIPSSQKQSFSEKS	1200
QY	1201	SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ	1260
DB	1201	SSGQSSKTEHMSSESSENTSTPSSNAKQNLHPSSAQSRGQPOKAATCKVSSINQETIQ	1260
QY	1261	TYCVEDTPICFSRCSLSSIAEDEIGCMTTOBADSANTLQIAEIKKIGTRSAEDPV	1320
DB	1261	TYCVEDTPICFSRCSLSSIAEDEIGCMTTOBADSANTLQIAEIKKIGTRSAEDPV	1320
QY	1321	SEVPAVSQHPHTKSRLOGSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPEHYVQET	1380
DB	1321	SEVPAVSQHPHTKSRLOGSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPEHYVQET	1380
QY	1381	PLMFSRCTSVSSLSDFESRSIASSVQSEPCSGMVGIIISPSDLPDPSGQTMPPSRKSTPP	1440
DB	1381	PLMFSRCTSVSSLSDFESRSIASSVQSEPCSGMVGIIISPSDLPDPSGQTMPPSRKSTPP	1440
QY	1441	PPQTAQTKRVPKNAKPTAEKRESGPKQAQVAVQVQVLPADTLLHFAFESTPDGF	1500
DB	1441	PPQTAQTKRVPKNAKPTAEKRESGPKQAQVAVQVQVLPADTLLHFAFESTPDGF	1500
QY	1501	SCSSLSALSALDEPFIQKVELRIMPVQENDNGNETESQPKESNENQEAETIDSE	1560
DB	1501	SCSSLSALSALDEPFIQKVELRIMPVQENDNGNETESQPKESNENQEAETIDSE	1560
QY	1561	KOLLDDSDDDIIELEECIISAMPTKSRKAKPAQATASKLPPPVARKPQOLPYKLLPS	1620
DB	1561	KOLLDDSDDDIIELEECIISAMPTKSRKAKPAQATASKLPPPVARKPQOLPYKLLPS	1620
QY	1621	QNRLOPKHVSFTPGDDMPRVYCVVEGTFINPSTATISLSDLTIESPPNELAAGEVGRGAQ	1680
DB	1621	QNRLOPKHVSFTPGDDMPRVYCVVEGTFINPSTATISLSDLTIESPPNELAAGEVGRGAQ	1680
QY	1681	SGEPEKRTIIEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINAMPKGKSHKP	1740
DB	1681	SGEPEKRTIIEGRSTDEAOGGKTSSVTIPELDDNKAEBGDIILAEICINAMPKGKSHKP	1740
QY	1741	FRVKKIMDOVOQASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNKL	1800
DB	1741	FRVKKIMDOVOQASASSAPNKNOLDGKKKPTSPVKPIPONTYRTRVRKNADSKNKL	1800
QY	1801	AERFVSDNKDSKQNLKNNKDFNDKLPNNEDVRGSAFDSPHHYTPIEGTPTCFGRND	1860
DB	1801	AERFVSDNKDSKQNLKNNKDFNDKLPNNEDVRGSAFDSPHHYTPIEGTPTCFGRND	1860
QY	1861	SLSLDLPDDDDVDLSREKAEELKAKENKESPAKVTSHTELTSNOQSANKTQAIKQPINR	1920
DB	1861	SLSLDLPDDDDVDLSREKAEELKAKENKESPAKVTSHTELTSNOQSANKTQAIKQPINR	1920
QY	1921	QQPKPILOKQSTFFQSSKDIIPRGAATDEKLNFAIENTPVCFSHNSLSLSLSDIDQENN	1980

Db 1921 GPKPILOKQSTFPQSSKIDPRGAATDEKLNFAIENTPVCFNSHSLSLSDIDQNN 1980
Qy 1981 NKENEPIKETEPPDQSGEPKQAGYAPKSFHVEDTPVCFNSHSLSLSDIDQNN 2040
Db 1981 NKENEPIKETEPPDQSGEPKQAGYAPKSFHVEDTPVCFNSHSLSLSDIDQNN 2040
Qy 2041 ECISAMPKXKXKPSRLKGDNEKHSRNMGGIIGEDLTLDKDIQRPDSEHGLSPDSENF 2100
Db 2041 ECISAMPKXKXKPSRLKGDNEKHSRNMGGIIGEDLTLDKDIQRPDSEHGLSPDSENF 2100
Qy 2101 WKAIQEGANSIYSSIHQAACLSRQASSDSDLSLKSGLISGSPHLPDQEKPT 2160
Db 2101 WKAIQEGANSIYSSIHQAACLSRQASSDSDLSLKSGLISGSPHLPDQEKPT 2160
Qy 2161 SNKGPRILKPGKSTLETKKIESEKGIKGGKVKVKSITGKVRNSISGOMKQLOAN 2220
Db 2161 SNKGPRILKPGKSTLETKKIESEKGIKGGKVKVKSITGKVRNSISGOMKQLOAN 2220
Qy 2221 MFSISGRFTMIHPGVNRNSSTSPVSKGPPKTPASKSPSEGTATTSPRGAKPSVKS 2280
Db 2221 MFSISGRFTMIHPGVNRNSSTSPVSKGPPKTPASKSPSEGTATTSPRGAKPSVKS 2280
Qy 2281 ELSPVARQTSQIGGSKAPSRGSDSTPSRPAQOPLSRPIOSGRNISPGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGGSKAPSRGSDSTPSRPAQOPLSRPIOSGRNISPGRNGISPPN 2340
Qy 2341 KLSQLPRTSSPTASTKSGSGKMYTSPGRMSQOQNTKQGLSKNASSIPRSESASK 2400
Db 2341 KLSQLPRTSSPTASTKSGSGKMYTSPGRMSQOQNTKQGLSKNASSIPRSESASK 2400
Qy 2401 LNMQNGNGANKVELSRMSSTKSGSDSDRERFVLVRQSTFIKEAPSPILRRKLESA 2460
Db 2401 LNMQNGNGANKVELSRMSSTKSGSDSDRERFVLVRQSTFIKEAPSPILRRKLESA 2460
Qy 2461 SFESLSPSSRPASPTRSQATPVLSPSLPDMSLTHSSVQAGGWRKLPNLSPTIYNDG 2520
Db 2461 SFESLSPSSRPASPTRSQATPVLSPSLPDMSLTHSSVQAGGWRKLPNLSPTIYNDG 2520
Qy 2521 RPAKRDHARSSESPRLPINRSGTWKREHSHSSLPVSTWRTGSSSILSASSES 2580
Db 2521 RPAKRDHARSSESPRLPINRSGTWKREHSHSSLPVSTWRTGSSSILSASSES 2580
Qy 2581 SEKASEDEKHVNSISGTKQSKENQVSAKGTWRKIKENEFPTNSTSTVSSGATNGAES 2640
Db 2581 SEKASEDEKHVNSISGTKQSKENQVSAKGTWRKIKENEFPTNSTSTVSSGATNGAES 2640
Qy 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNPVIDSVSEKAMPNIKDSKDN 2700
Db 2641 KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNPVIDSVSEKAMPNIKDSKDN 2700
Qy 2701 QAKQNVGSGVPMRTVGLNRLNSFIQDAPQKTEIKPGQNNPVVSETNESSIVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLNSFIQDAPQKTEIKPGQNNPVVSETNESSIVERT 2760
Qy 2761 PFSSSSSKHSSPSGTVAARVTPFNYPNPSRKSSADTSARPSQIPTPVNNTKKRDSKT 2820
Db 2761 PFSSSSSKHSSPSGTVAARVTPFNYPNPSRKSSADTSARPSQIPTPVNNTKKRDSKT 2820
Qy 2821 DSTESSGTQSPKRRHSGSYLVTSV 2843
Db 2821 DSTESSGTQSPKRRHSGSYLVTSV 2843

RESULT 2
AAW76144
ID AAW76144 standard; protein; 2843 AA.
XX
AC AAW76144;
XX
DT 25-MAR-2003 (revised)
DT 23-NOV-1998 (first entry)
XX

Human APC protein #2.
XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
XX chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour; FAP;
XX Gardner's Syndrome; GS; predisposition.
OS Homo sapiens.
XX US5783666-A.
XX 21-JUL-1998.
XX 25-MAY-1995; 95US-00452655.
XX 16-JAN-1991; 91GB-00000962.
XX 16-JAN-1991; 91GB-00000963.
XX 16-JAN-1991; 91GB-00000974.
XX 16-JAN-1991; 91GB-00000975.
XX 08-AUG-1991; 91US-00741940.
XX 12-AUG-1994; 94US-00289548.
XX (CANC-) CANCER INST.
XX (UYJO) UNIV JOHNS HOPKINS.
XX (UTAH) UNIV UTAH.
XX (ZENE) ZENEC PHARM.
XX Kinzler K, Joslyn G, Markham AF, Carlson M, White RL;
XX Thliveris A, Groden J, Anand R, Nakamura Y, Albertsen H;
XX Vogelstein B, Hedge PJ;
XX WPI; 1998-427100/36.
XX Adenomatous polyposis coli protein - useful in the treatment of cancers
XX associated with mutation(s) on human chromosome 5q21.
XX Disclosure; Col 63-78; 102pp; English.
XX This sequence represents a human familial adenomatous polyposis coli
XX (APC) protein isolated from 87 cDNA clones. The gene for the protein is
XX present on human chromosome 5q21 and is also referred to as adenomatous
XX polyposis coli gene. It is a tumour suppressor gene, and mutations in
XX this gene have been associated with tumorigenesis in retinoblastoma and
XX colorectal tumours, and especially familial adenomatous polyposis (FAP)
XX and Gardner's Syndrome (GS). The protein can be used in therapy to
XX replace lack of native functional protein and the nucleic acids can be
XX used for gene therapy. The nucleic acids that encode them can also be
XX used as probes and primers in detection of the cancers and predisposition
XX to it. (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 2843 AA;
Query Match 100.0%; Score 14566; DB 2; Length 2843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAAASYDQLLKQVEALKWENSLRQLEDNSNHLTKLETEASNKVEVLKQLQGSIEDEAM 60
Db 1 MAAASYDQLLKQVEALKWENSLRQLEDNSNHLTKLETEASNKVEVLKQLQGSIEDEAM 60
Qy 61 ASSQIDLLRLKELNLDSSNFPQVKLRISKMSLRSGSREGSVSSRSGECPVPMGSPFR 120
Db 61 ASSQIDLLRLKELNLDSSNFPQVKLRISKMSLRSGSREGSVSSRSGECPVPMGSPFR 120
Qy 121 RGFVNGSRESTGYLEELKERSLLADLDKEEKKDYVAQLQNLTKRIDSLPTNFSL 180
Db 121 RGFVNGSRESTGYLEELKERSLLADLDKEEKKDYVAQLQNLTKRIDSLPTNFSL 180
Qy 181 QTMTRQLEYEARQIRVAMEEQIGTCQDMKEKQAQRRIARIQIQIEKDIIRIQLQSQAT 240
Db 181 QTMTRQLEYEARQIRVAMEEQIGTCQDMKEKQAQRRIARIQIQIEKDIIRIQLQSQAT 240
Qy 241 EABESSONKHETGSHDAERONEGOGVEINWATSGNGQSTTMDHETASVSSSSHTSA 300

Db 241 EAERSSQKHETGSHDAERQNEGGVGEINMATSNGQGSGTTEMHDHETASVLSSSTHSA 300
Qy 301 PRRLTSHLGTKEVMVYSLLSMLGTHDKDDMSRRTLAMSSQSDSCIWMROSGLPLLIQLL 360
Db 301 PRRLTSHLGTKEVMVYSLLSMLGTHDKDDMSRRTLAMSSQSDSCIWMROSGLPLLIQLL 360
Qy 361 HGNDKOSVLGNRSGSKEARASAAALHNLIIHSQDDKGRREIRVLHLEQIRAYCETC 420
Db 361 HGNDKOSVLGNRSGSKEARASAAALHNLIIHSQDDKGRREIRVLHLEQIRAYCETC 420
Qy 421 WEMQEAHEPGMDQDNPMAPVPEHQICPAVCVLMKLSFDEHRHAMNELGGLQAI AELLQ 480
Db 421 WEMQEAHEPGMDQDNPMAPVPEHQICPAVCVLMKLSFDEHRHAMNELGGLQAI AELLQ 480
Qy 481 VDCBMYGLTNDHYSITLRRYAGMALTNLTGDIVANKATILCSMGCMRALVAQLKSEEDL 540
Db 481 VDCBMYGLTNDHYSITLRRYAGMALTNLTGDIVANKATILCSMGCMRALVAQLKSEEDL 540
Qy 541 QOVIASVLRLNLRADVNSKKTIREVGSVKALMECALEVKKESTLKSLSALWNL SAHCT 600
Db 541 QOVIASVLRLNLRADVNSKKTIREVGSVKALMECALEVKKESTLKSLSALWNL SAHCT 600
Qy 601 ENKADI CAVDGAALAFVLTGTYRSQTNWTLAIIESGGGILNVSSLTATNEDHROI LRENN 660
Db 601 ENKADI CAVDGAALAFVLTGTYRSQTNWTLAIIESGGGILNVSSLTATNEDHROI LRENN 660
Qy 661 CLOTLLOHLKSHSLTIVSNACGTLWNL SARPKDQEALWDMGAVSMLKNLIHSKHWMIAM 720
Db 661 CLOTLLOHLKSHSLTIVSNACGTLWNL SARPKDQEALWDMGAVSMLKNLIHSKHWMIAM 720
Qy 721 GSAAALRLMANRPAKYKDANIMSPGSLPSLHV RKQKALEABEADAOHSETFDNIDNLS 780
Db 721 GSAAALRLMANRPAKYKDANIMSPGSLPSLHV RKQKALEABEADAOHSETFDNIDNLS 780
Qy 781 PKASHRSKORHKQSLG DYVFDTRHDDNDSDFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
Db 781 PKASHRSKORHKQSLG DYVFDTRHDDNDSDFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
Qy 841 LDSRSSEKRSLEBERGIGLGNYPATENFGTSKRGLOISTTAAQIAKYMEVSAIHTS 900
Db 841 LDSRSSEKRSLEBERGIGLGNYPATENFGTSKRGLOISTTAAQIAKYMEVSAIHTS 900
Qy 901 QEDRSSGSTTELHCVTDERNALRSSAAHSTNFTNFKSENSNRITCSMPYAKLEYKRSS 960
Db 901 QEDRSSGSTTELHCVTDERNALRSSAAHSTNFTNFKSENSNRITCSMPYAKLEYKRSS 960
Qy 961 NDSLNSVSSSDGYKRGOMKPSIESYSEDDESKFCSYGQYPADLAHKIHSANHMDNDGE 1020
Db 961 NDSLNSVSSSDGYKRGOMKPSIESYSEDDESKFCSYGQYPADLAHKIHSANHMDNDGE 1020
Qy 1021 LDTPTINTSLKYSDQLNSGRQSPQNERWARPKHII EDEIKOSEQRQSRNQSTTYPVYTE 1080
Db 1021 LDTPTINTSLKYSDQLNSGRQSPQNERWARPKHII EDEIKOSEQRQSRNQSTTYPVYTE 1080
Qy 1081 STDDKHLKFQPHFGQECVSPYRSGRANGSETNRVGNHGINQVNSQSLCEQEDDYEDDKP 1140
Db 1081 STDDKHLKFQPHFGQECVSPYRSGRANGSETNRVGNHGINQVNSQSLCEQEDDYEDDKP 1140
Qy 1141 TNYSERVSEEBQHEEBERPTNYSIKYNEKHXVDQPIDYSLKYATDIPSSQKQSFSPSKS 1200
Db 1141 TNYSERVSEEBQHEEBERPTNYSIKYNEKHXVDQPIDYSLKYATDIPSSQKQSFSPSKS 1200
Qy 1201 SSGOSSKTEHMSSENSTSPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQETIQ 1260
Db 1201 SSGOSSKTEHMSSENSTSPSSNAKQNLHPSSAQRSGQPOKAATCKVSSINQETIQ 1260
Qy 1261 TYCVEDTPICFSRCSLSLSLSSADEICGNQTTQEADSANTLQIAEIKEKIGTRSAEDPV 1320
Db 1261 TYCVEDTPICFSRCSLSLSLSSADEICGNQTTQEADSANTLQIAEIKEKIGTRSAEDPV 1320
Qy 1321 SEVPAVSQHPRTKSRLOGSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPEHYVQET 1380
Db 1321 SEVPAVSQHPRTKSRLOGSSLSSESARHKAVERSSGAKSPSKGAOTPKSPPEHYVQET 1380

Qy 1381 PLMFSRCTSVSSLDSPESRSRTASSVQSEPCSGMVGGIISPSDLPSDQTMPPRSRKTTP 1440
Db 1381 PLMFSRCTSVSSLDSPESRSRTASSVQSEPCSGMVGGIISPSDLPSDQTMPPRSRKTTP 1440
Qy 1441 PPPQTAQTKREVVPKNAKPTAEKRESGPKQAAVNAAVQVQLPDADTLHLHFATESTPDGF 1500
Db 1441 PPPQTAQTKREVVPKNAKPTAEKRESGPKQAAVNAAVQVQLPDADTLHLHFATESTPDGF 1500
Qy 1501 SCSSLSLSALSDDEPFIQKVELRIMPVQENDNGNETESQPKESNENQKEAEKTTDSE 1560
Db 1501 SCSSLSLSALSDDEPFIQKVELRIMPVQENDNGNETESQPKESNENQKEAEKTTDSE 1560
Qy 1561 KOLLDDSDDDDIETLEECIIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Db 1561 KOLLDDSDDDDIETLEECIIISAMPTKSRKAKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
Qy 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLDTIESPPNELAAGEVRGAQ 1680
Db 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLDTIESPPNELAAGEVRGAQ 1680
Qy 1681 SGEPEKRTDITPEGRSTDEAOGGKTSSVTIPELDDNKAEBGDI LAECINSAMPKKGSHKP 1740
Db 1681 SGEPEKRTDITPEGRSTDEAOGGKTSSVTIPELDDNKAEBGDI LAECINSAMPKKGSHKP 1740
Qy 1741 FRVKKIMDQVOQASASSAPNKNQIDGKKKPTSPVXPIQONTYRTRVRKNADSKNIN 1800
Db 1741 FRVKKIMDQVOQASASSAPNKNQIDGKKKPTSPVXPIQONTYRTRVRKNADSKNIN 1800
Qy 1801 AERVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVRGSPAFDPSPHYTPIECTPYCFSRND 1860
Db 1801 AERVFSDNKKSKQNLKNNSKDFNDKLPNNEDRVRGSPAFDPSPHYTPIECTPYCFSRND 1860
Qy 1861 SLSLSDFDDDDDVLSREKAELRKAKENKESBAKVTSHTELTSNQOSANKTOAI AKQPINR 1920
Db 1861 SLSLSDFDDDDDVLSREKAELRKAKENKESBAKVTSHTELTSNQOSANKTOAI AKQPINR 1920
Qy 1921 GQPKILOQKSTFPQSSKDIIDRGAATDEKLOFAIENTPVCFSHNSSLSSLSIDIDENN 1980
Db 1921 GQPKILOQKSTFPQSSKDIIDRGAATDEKLOFAIENTPVCFSHNSSLSSLSIDIDENN 1980
Qy 1981 NKENEPIKETETPPDPSQGEPSKPQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDLLQ 2040
Db 1981 NKENEPIKETETPPDPSQGEPSKPQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDLLQ 2040
Qy 2041 ECISISSAMPKKKPSRLKGDNEKHSFRNMGGITLGEDLTLDLKDIORPDEHGLSPDSNFD 2100
Db 2041 ECISISSAMPKKKPSRLKGDNEKHSFRNMGGITLGEDLTLDLKDIORPDEHGLSPDSNFD 2100
Qy 2101 WKAIQEGANSIVSSLHQAAAAACL SRQASSSDSISLKSGLSGSPHLLTPDQEEKPT 2160
Db 2101 WKAIQEGANSIVSSLHQAAAAACL SRQASSSDSISLKSGLSGSPHLLTPDQEEKPT 2160
Qy 2161 SNKGPRILLKPEKSTLETTKTIESSEKGIKGGKVYKSLITOKVRSNBSISQMKQPIQAN 2220
Db 2161 SNKGPRILLKPEKSTLETTKTIESSEKGIKGGKVYKSLITOKVRSNBSISQMKQPIQAN 2220
Qy 2221 MPISIRGRMTIHIPOVRNSSSTSPVSKGPPKLTTPASKSPSEGTATTSRPGAKPSVKS 2280
Db 2221 MPISIRGRMTIHIPOVRNSSSTSPVSKGPPKLTTPASKSPSEGTATTSRPGAKPSVKS 2280
Qy 2281 ELSPVARQTSQIGSSKAPSRSGRSDSTPSRPAQOPLSRPICSQGRNBSISQGRNGISPPN 2340
Db 2281 ELSPVARQTSQIGSSKAPSRSGRSDSTPSRPAQOPLSRPICSQGRNBSISQGRNGISPPN 2340
Qy 2341 KLSQIPRTSSPSTASTKSSGSKWYSYTPGQMSQOQLTKOTGLSKNASSIPRSESASKG 2400
Db 2341 KLSQIPRTSSPSTASTKSSGSKWYSYTPGQMSQOQLTKOTGLSKNASSIPRSESASKG 2400
Qy 2401 LNQMNGANGANKVELSRMSSTKSSGESDESRSERFVLVRQSTFIKEAPSTLRRLKEESA 2460
Db 2401 LNQMNGANGANKVELSRMSSTKSSGESDESRSERFVLVRQSTFIKEAPSTLRRLKEESA 2460

QY 2461 SFESLSPSRPASPTSRQATPVLSPLPDMSLSTHSSVQAGWRLPPNLSPTTIEYNDG 2520
Db 2461 SFESLSPSRPASPTSRQATPVLSPLPDMSLSTHSSVQAGWRLPPNLSPTTIEYNDG 2520
QY 2521 RPAKRHDIAARSHSESRLPPINRSGTWKREHSHSSSLPRVSTWRTGSSSILGASSES 2580
Db 2521 RPAKRHDIAARSHSESRLPPINRSGTWKREHSHSSSLPRVSTWRTGSSSILGASSES 2580
QY 2581 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSTNSTQTVSSGATNGAES 2640
Db 2581 SEKAKSEDEKHVNSISGTSQSKENQVSAKGTWRKIKENEFSTNSTQTVSSGATNGAES 2640
QY 2641 KTLIYQMAFAVAKTDEVDVWRIEDCPINPRSGRSPGTGNTPPVIDSVSEKANPNIKSDXDN 2700
Db 2641 KTLIYQMAFAVAKTDEVDVWRIEDCPINPRSGRSPGTGNTPPVIDSVSEKANPNIKSDXDN 2700
QY 2701 QAKQNVGSGVPMRTVGLNRLNSFTQVDAPOKGTETKPGQNPVPVSETNESSIVERT 2760
Db 2701 QAKQNVGSGVPMRTVGLNRLNSFTQVDAPOKGTETKPGQNPVPVSETNESSIVERT 2760
QY 2761 PFSSSSSKHSPSGTVAARVTFPNYNPSPKSSADSTSARPSQIPTPVNNNTKRDGKT 2820
Db 2761 PFSSSSSKHSPSGTVAARVTFPNYNPSPKSSADSTSARPSQIPTPVNNNTKRDGKT 2820
QY 2821 DSTESGTSQPKRHSGSYLVTGV 2843
Db 2821 DSTESGTSQPKRHSGSYLVTGV 2843

RESULT 3
ID AAB23011 standard; protein; 2843 AA.
XX AAB23011;
AC AAB23011;
DT 16-JAN-2001 (first entry)
XX Human APC protein (splice variant 1).
DE
XX APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;
KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;
KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;
KW bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;
KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;
KW genetic predisposition; drug screening; DP2.5; splice variant.
XX
OS Homo sapiens.
XX
XX US6114124-A.
XX
PD 05-SEP-2000.
XX
PF 25-MAY-1995; 95US-00450582.
XX
PR 16-JAN-1991; 91GB-000000962.
PR 16-JAN-1991; 91GB-000000963.
PR 16-JAN-1991; 91GB-000000974.
PR 16-JAN-1991; 91GB-000000975.
PR 08-AUG-1991; 91US-00741940.
PR 12-AUG-1994; 94US-00289548.
XX
PA (ICIL) IMPERIAL CHEM IND PLC.
PA (UYGO) UNIV JOHNS HOPKINS.
PA (UTAH) UNIV UTAH.
PA (CANC-) CANCER INST.
XX
PI Carlson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R;
PI Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B;
PI Hedge PJ;
XX
DR WPI; 2000-565003/52.
DR N-PSDB; AAA93449.
XX

Detecting Adenomatous Polyposis Coli (APC) protein in a sample for
diagnosing cancers, involves contacting the sample with antibodies that
specifically bind to APC protein and detecting the complex formed.
Claim 1; Fig 3A-C; 125pp; English.
The invention relates to a novel method for detecting Adenomatous
Polyposis Coli (APC) protein in a sample. The method involves contacting
the sample with antibodies which specifically binds to the 2843 amino
acid form of the human APC protein, or to a mutant APC protein, and
detecting an APC-antibody complex. Mutations in the APC gene play a role
in tumorigenesis, indicating that it is a tumour suppressor gene. It is
located on chromosome 5q21, which corresponds to the FAP (familial
adenomatous polyposis) locus. FAP is an autosomal dominant inherited
disease in which affected individuals develop hundreds to thousands of
adenomatous polyps in the colon and rectum, some of which progress to
malignancy. The FAP locus is often found to be deleted in sporadic (i.e.,
non-familial) adenomas and carcinomas, and chromosome 5q deletions have
also been observed in tumours of the lung, breast, colon, rectum,
bladder, liver, sarcoma, stomach, and prostate, and in leukaemias and
lymphomas. Although the FAP locus contains several other genes such as
FER, TBL1, TB2, and MCC, it is thought that mutations in the APC gene play
a key role in the development of FAP and sporadic tumours. The method is
useful for detecting APC protein and its mutant forms in foetal tissue,
placental tissue, amniotic fluid, blood, serum or a tumour sample. The
method is useful for diagnosing or prognosing neoplastic tissue, for
detecting a genetic predisposition to cancer, for detecting germline and
somatic alteration of wild-type APC genes, and for testing therapeutic
agents for the ability to suppress tumours. The present sequence
represents a 2843 amino acid splice variant of the human APC protein.
This variant is more abundant than the 2742 amino acid variant (AAB23012)
Sequence 2843 AA;

Query Match 100.0%; Score 14566; DB 3; Length 2843;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAASVYDQLLKQVEALKWENSLRQLEDNSNHLTKLETSANMKVYLKQSGSIEDEAM 60
Db 1 MAASVYDQLLKQVEALKWENSLRQLEDNSNHLTKLETSANMKVYLKQSGSIEDEAM 60
QY 61 ASSQIDLLRLKELNLDSSNFFGVKLRSMVRSYSGREGSVSSRSGEGSPVPMGSPFR 120
Db 61 ASSQIDLLRLKELNLDSSNFFGVKLRSMVRSYSGREGSVSSRSGEGSPVPMGSPFR 120
QY 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKWYVLAQLNLTKRIDSLPTENFSL 180
Db 121 RGFVNGSRESTGYLEELKERSILLADLDKEEKEKWYVLAQLNLTKRIDSLPTENFSL 180
QY 181 QTDWTRRQLEYEARQIRVAMEBOJGTCQDMKEKRAQRIARIQIQIEKDILRIQLLOSQAT 240
Db 181 QTDWTRRQLEYEARQIRVAMEBOJGTCQDMKEKRAQRIARIQIQIEKDILRIQLLOSQAT 240
QY 241 EAERSSONKHETGSHDAERONEGGVGEINWATSGNGQSTTMDHETASVLSSTHSA 300
Db 241 EAERSSONKHETGSHDAERONEGGVGEINWATSGNGQSTTMDHETASVLSSTHSA 300
QY 301 PRRLTSHLGTKEVMYVLSLMLGTHDKDDMSRLLTAMSSSSQSDSCISNRSGCPLLIQLL 360
Db 301 PRRLTSHLGTKEVMYVLSLMLGTHDKDDMSRLLTAMSSSSQSDSCISNRSGCPLLIQLL 360
QY 361 HGNDKDSVLLGNRSGSKEAPARASALHNIISQPDCKRGRRIIRVHLLLEQIRAYCETC 420
Db 361 HGNDKDSVLLGNRSGSKEAPARASALHNIISQPDCKRGRRIIRVHLLLEQIRAYCETC 420
QY 421 WEQEAHEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQIAIELQ 480
Db 421 WEQEAHEPGMDQDKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGQIAIELQ 480
QY 481 VDCMYGLTNDHYSITLRRYAGMALNLTFGDVANKATLCMSKGMCRALVAQLKSESEDL 540
Db 481 VDCMYGLTNDHYSITLRRYAGMALNLTFGDVANKATLCMSKGMCRALVAQLKSESEDL 540

541 QOVIASVLRNLSWRADVNSKTLREYGSVKALMECALEVKESTLKSVLASLWNLSAHCT 600
 541 QOVIASVLRNLSWRADVNSKTLREYGSVKALMECALEVKESTLKSVLASLWNLSAHCT 600
 601 ENKADI CAVDGALAFVLGTLTYRSQNTWLAIESGGGILRNVSLLIATNEDHQILRENN 660
 601 ENKADI CAVDGALAFVLGTLTYRSQNTWLAIESGGGILRNVSLLIATNEDHQILRENN 660
 661 CLQTLQHLKSHSLTIVSNACGLTWNLSARNPKDQALWDMGAVSMKXLIHSKXWMIAM 720
 661 CLQTLQHLKSHSLTIVSNACGLTWNLSARNPKDQALWDMGAVSMKXLIHSKXWMIAM 720
 721 GSAALRNLMANRPAYKDANIMSPGSSIPSLHVRKQKALEAELDAQHLSSETFDNIDNLS 780
 721 GSAALRNLMANRPAYKDANIMSPGSSIPSLHVRKQKALEAELDAQHLSSETFDNIDNLS 780
 781 PKASHRSKQHKOSLYGDIYVDFNRHDDNPNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
 781 PKASHRSKQHKOSLYGDIYVDFNRHDDNPNFNTGNMTVLSPLYLNTTVLPSSSSSRGS 840
 841 LDSRSSEKDRSLRERIGIGLNYHPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTS 900
 841 LDSRSSEKDRSLRERIGIGLNYHPATENPGTSSKRGLOISTTAAQIAKVMEEVSAIHTS 900
 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHSNYNTFTKSENSNRTCSPYAKLEYKRS 960
 901 QEDRSSGTTTELHCVTDERNALRRSSAAHSHSNYNTFTKSENSNRTCSPYAKLEYKRS 960
 961 NDSLSNVSSSDGCGKQOMKPSIESYSEDDDESFCYGOYPADLAHKIHSANHMDNDGE 1020
 961 NDSLSNVSSSDGCGKQOMKPSIESYSEDDDESFCYGOYPADLAHKIHSANHMDNDGE 1020
 1021 LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRNQSTTYPVYTE 1080
 1021 LDTPINYSKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRNQSTTYPVYTE 1080
 1081 STDDKHLLKQPHFQOCVCPYRSRGANGSETNVRVGNHGINQVOSLQCEDDYEDDKP 1140
 1081 STDDKHLLKQPHFQOCVCPYRSRGANGSETNVRVGNHGINQVOSLQCEDDYEDDKP 1140
 1141 TNYSERYSSEEEHQBEEERPTNYSIKYNEKRHVDPQIDYSLKYATDIPSSQKQSPFSKS 1200
 1141 TNYSERYSSEEEHQBEEERPTNYSIKYNEKRHVDPQIDYSLKYATDIPSSQKQSPFSKS 1200
 1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAQSRSQPKAATCKVSSINQETIQ 1260
 1201 SSGOSSKTEHMSSENSTPSSNAKQNLHPSSAQSRSQPKAATCKVSSINQETIQ 1260
 1261 TYCVEDTPICFSRCSLSLSAEDGICQNTQOADSANTLOIAETKEKIGTRSAEDPV 1320
 1261 TYCVEDTPICFSRCSLSLSAEDGICQNTQOADSANTLOIAETKEKIGTRSAEDPV 1320
 1321 SEVPAVSQHPRTKSRIOGSSLSSEARHKAVERSSGAKSPSKGAOTPKSPHEHYVOET 1380
 1321 SEVPAVSQHPRTKSRIOGSSLSSEARHKAVERSSGAKSPSKGAOTPKSPHEHYVOET 1380
 1381 PLMFSRCTSVSSLDSPFSRSFASVQSEPCGMVSGIISPSDLDPSPGQWMPFSRSKTPP 1440
 1381 PLMFSRCTSVSSLDSPFSRSFASVQSEPCGMVSGIISPSDLDPSPGQWMPFSRSKTPP 1440
 1441 PPPQTAQKREVPKNKAPTAKEKREGPKQAANVAORVQVLPDADTLHLHATESPDGF 1500
 1441 PPPQTAQKREVPKNKAPTAKEKREGPKQAANVAORVQVLPDADTLHLHATESPDGF 1500
 1501 SCSSLSALSJDEEPIQKQVLRIMPVQENDNGNETESEOPEKSENQOEKAEKTIIDSE 1560
 1501 SCSSLSALSJDEEPIQKQVLRIMPVQENDNGNETESEOPEKSENQOEKAEKTIIDSE 1560
 1561 KDLLDDSDDDZIEILEECIIISAMPTKSRKAKQAQATASKLPPPVARKPQOLPVYKLLPS 1620
 1561 KDLLDDSDDDZIEILEECIIISAMPTKSRKAKQAQATASKLPPPVARKPQOLPVYKLLPS 1620

1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLDTIESPNELAAAGVGRRGAQ 1680
 1621 QNRLOPQKHVSFTPGDDMPRVYCVEGTPIINFSTATISLDTIESPNELAAAGVGRRGAQ 1680
 1681 SGEPEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAEICINAMPKGSHPK 1740
 1681 SGEPEKEDTTPTEGRSTDEAQQGKTSSVTIPELDDNKABEGDILAEICINAMPKGSHPK 1740
 1741 FRVKIMDOVQOASASASAPNKQLDGKKKPTSPVKPIPONTTEYTRVRKNADSKNNLN 1800
 1741 FRVKIMDOVQOASASASAPNKQLDGKKKPTSPVKPIPONTTEYTRVRKNADSKNNLN 1800
 1801 AERVFNSKDKSKONLKNNKDFNDKLPNNEDVRGSAFDSPHHTPIEBGTPCYFSRND 1860
 1801 AERVFNSKDKSKONLKNNKDFNDKLPNNEDVRGSAFDSPHHTPIEBGTPCYFSRND 1860
 1861 SLSLDFDDDDVDLSREKAEIRKAKENKSEAKVTSHTELTSNQOQANKTOAKQPINR 1920
 1861 SLSLDFDDDDVDLSREKAEIRKAKENKSEAKVTSHTELTSNQOQANKTOAKQPINR 1920
 1921 QCPKPILOKQSTFPQSSKDIPTDRGAATDEKLQNPATIENTPVCFSHNSLSLSDIDQENN 1980
 1921 QCPKPILOKQSTFPQSSKDIPTDRGAATDEKLQNPATIENTPVCFSHNSLSLSDIDQENN 1980
 1981 NKENEPKETEPDPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
 1981 NKENEPKETEPDPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNSSLSLSIDSEDDLLQ 2040
 2041 ECISAMPKPKKRLKGDNEKHSRNMGGITLGDLDLKDQORPDSEHGLSPDSNED 2100
 2041 ECISAMPKPKKRLKGDNEKHSRNMGGITLGDLDLKDQORPDSEHGLSPDSNED 2100
 2101 WKATOEGANSIVSLHQAAAAACLSRQASDSDSILSKSGISLGSFPHLTPDOEKPF 2160
 2101 WKATOEGANSIVSLHQAAAAACLSRQASDSDSILSKSGISLGSFPHLTPDOEKPF 2160
 2161 SNKGPRIILKPEKSTLETTKIETESKGIKGGKVKYKSLITGKVRNSSEISGOMKQPIQAN 2220
 2161 SNKGPRIILKPEKSTLETTKIETESKGIKGGKVKYKSLITGKVRNSSEISGOMKQPIQAN 2220
 2221 MPSISRGTMTHIPGVNRNSSSTSPVSKGPKPLKTPAKSPSEGTATTPSPGAKPSVKS 2280
 2221 MPSISRGTMTHIPGVNRNSSSTSPVSKGPKPLKTPAKSPSEGTATTPSPGAKPSVKS 2280
 2281 ELSPVARTSIOIGSSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISFGRNGISPPN 2340
 2281 ELSPVARTSIOIGSSSKAPSRSGSRDSTPSRPAQOPLSRPIQSPGRNSISFGRNGISPPN 2340
 2341 KLSQIPRTSSPSTASTKSSGSKYKSYTSPGRMSQOQNLTKOTGLSKQASSIPRSESASKG 2400
 2341 KLSQIPRTSSPSTASTKSSGSKYKSYTSPGRMSQOQNLTKOTGLSKQASSIPRSESASKG 2400
 2401 LQNMNNGANGANKVLSRYMSYTSKSGSEDSRSEPVLRQSTFTKEAPSPTRKLEESA 2460
 2401 LQNMNNGANGANKVLSRYMSYTSKSGSEDSRSEPVLRQSTFTKEAPSPTRKLEESA 2460
 2461 SPESISLSPSRPASPRSQAOQFVLSPSLPDMSLSTHSSVQAGWKLPNLSPIETENDG 2520
 2461 SPESISLSPSRPASPRSQAOQFVLSPSLPDMSLSTHSSVQAGWKLPNLSPIETENDG 2520
 2521 RPAKHDIARSHSESFRSLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 2521 RPAKHDIARSHSESFRSLPINRSGTWKREHSHSSSLPRVSTWRTGSSSSILSASSES 2580
 2581 SEKAKSEDEKHVNSISGTSKQKNOVSAGTWRKIKENEFPTNSTQTSVSGATNGAES 2640
 2581 SEKAKSEDEKHVNSISGTSKQKNOVSAGTWRKIKENEFPTNSTQTSVSGATNGAES 2640
 2641 KTLIIQMAPAVSKTEDVWVRIEDCFINNPRSGRSPGTGNTPPVIDSVSEKANPNIKSDKN 2700
 2641 KTLIIQMAPAVSKTEDVWVRIEDCFINNPRSGRSPGTGNTPPVIDSVSEKANPNIKSDKN 2700
 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVADPDQKTEIKPGQNNPVVSEINNESSIVERT 2760

Db 2701 CAQNVGVGSPMRTVGLNEBLSIQVDAPQKGTETKPGQNPVPVSETNESSIVERT 2760
 QY 2761 PFSSSSSKHSPSGTVAARTVPPFNYPSPKSSADSTSARPSQIPTPPVNNTKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARTVPPFNYPSPKSSADSTSARPSQIPTPPVNNTKRDSKT 2820
 QY 2821 DSTESSGTQSPKRHSGSYLVTSV 2843
 Db 2821 DSTESSGTQSPKRHSGSYLVTSV 2843
 RESULT 4
 ABG71105
 ID ABG71105 standard; protein; 2843 AA.
 AC ABG71105;
 DT 08-JAN-2003 (first entry)
 DE Human adenomatous polyposis coli (APC) protein.
 XX Human; adenomatous polyposis coli; APC; transgenic animal; colon cancer;
 KW developmental abnormality; anticancer drug; beta-catenin;
 KW Wnt/Wg signalling pathway.
 XX Homo sapiens.
 OS Homo sapiens.
 PN EPI243646-A2.
 FN 25-SEP-2002.
 PD 07-MAR-2002; 2002EP-00290571.
 PF 19-MAR-2001; 2001US-0276483P.
 PR (COUL) COUNCIL SCI & IND RES.
 FA Bhandari P, Shashidhara LS;
 PI WPI; 2002-742695/81.
 DR New transgenic Drosophila containing the human colon cancer gene
 PT Adenomatous Polyposis Coli is useful as an assay model to screen for new
 PT drugs, particularly against colon cancer.
 XX Claim 1; Page 19-20; 52pp; English.
 CC The invention describes a transgenic Drosophila whose genome comprises
 CC the full-length human colon cancer gene Adenomatous Polyposis Coli (APC)
 CC allowing regulated mis-expression of the APC gene resulting in
 CC developmental abnormalities. The transgenic flies are used to screen and
 CC validate efficacy of anticancer drugs, to identify new target proteins
 CC interacting with beta-catenin, genes which interact with human APC, to
 CC study the biochemical function of human APC and to identify additional
 CC components of the Drosophila Wnt/Wg signalling pathway. In particular the
 CC flies are used to screen potential drugs against colon cancer. This is
 CC the amino acid sequence of the human adenomatous polyposis coli (APC)
 CC protein DNA encoding which is incorporated in the transgenic flies
 XX Sequence 2843 AA;
 SQ
 Query Match 100.0%; Score 14566; DB 5; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAASYDQLLKQVEALKMNSNLRQLEDNHNLTLEASNMKEVLKQLQGSIEDEAM 60
 Db 1 MAAASYDQLLKQVEALKMNSNLRQLEDNHNLTLEASNMKEVLKQLQGSIEDEAM 60
 QY 61 ASSGQIDLLERLKEINLDSNFPFGVKLRKSKMSLRSGREGSVSRSGECSFVPMGSPFR 120
 Db 61 ASSGQIDLLERLKEINLDSNFPFGVKLRKSKMSLRSGREGSVSRSGECSFVPMGSPFR 120

QY 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEKDWYVAQLQNLTKRIDSLPTNFSL 180
 Db 121 RGFVNGSRSTGYLELEKERSLLADLDKEEKEKDWYVAQLQNLTKRIDSLPTNFSL 180
 QY 181 QTMTRRQLEYEARQIRVAMEEQLTGTCQDWEKRAQRRIARIQOIEKDIIRIROLLOSOAT 240
 Db 181 QTMTRRQLEYEARQIRVAMEEQLTGTCQDWEKRAQRRIARIQOIEKDIIRIROLLOSOAT 240
 QY 241 EAERSSQNHETGSHDAERQNEQGVGEINMATSNGQSGSTTMDHETASVLSSSSTHSA 300
 Db 241 EAERSSQNHETGSHDAERQNEQGVGEINMATSNGQSGSTTMDHETASVLSSSSTHSA 300
 QY 301 PRELTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQSCISMROGCLPLLIQLL 360
 Db 301 PRELTSHLGTKEVMYSLLSMLGTHDKDMSRTLLAMSSQSCISMROGCLPLLIQLL 360
 QY 361 HGNDKDSVLLGNSRGSKEARASAAALHNIISQPDDEKRRREIRVHLLEQIRAYCETC 420
 Db 361 HGNDKDSVLLGNSRGSKEARASAAALHNIISQPDDEKRRREIRVHLLEQIRAYCETC 420
 QY 421 WEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLOAIELLQ 480
 Db 421 WEQEAHEPGMDQKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGLOAIELLQ 480
 QY 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTFDGVANKATILCSMKGCMRALVAQLKSSIDL 540
 Db 481 VDCEMYGLTNDHYSITLRRYAGMALTNLTFDGVANKATILCSMKGCMRALVAQLKSSIDL 540
 QY 541 QQVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSLSALWNLSAHCT 600
 Db 541 QQVIASVLRNLSWRADVNSKKTIREVGSVKALMECALEVKESTLKSLSALWNLSAHCT 600
 QY 601 ENKADICAVDGAFLVGLTYRSQNTLAIIESGGIILRNVSLLIATNEDHROIIRENN 660
 Db 601 ENKADICAVDGAFLVGLTYRSQNTLAIIESGGIILRNVSLLIATNEDHROIIRENN 660
 QY 661 CLQTLQHLKSHSLTIVSNACGLTNLSARNPKQDALWDMGAVSMKLIHSKHWMIAM 720
 Db 661 CLQTLQHLKSHSLTIVSNACGLTNLSARNPKQDALWDMGAVSMKLIHSKHWMIAM 720
 QY 721 GSAALRLNLMANRPKAKIDANIMSPGSLPSLHVRKQKALEAELDAQHLSETFDNIDNLS 780
 Db 721 GSAALRLNLMANRPKAKIDANIMSPGSLPSLHVRKQKALEAELDAQHLSETFDNIDNLS 780
 QY 781 PKASHRSKQHKOSLYGDIYVDTNRHDDNRSNDFNTGNMTVLSPLYNTTVLPSSSSRGS 840
 Db 781 PKASHRSKQHKOSLYGDIYVDTNRHDDNRSNDFNTGNMTVLSPLYNTTVLPSSSSRGS 840
 QY 841 LDSRSSEKORSLEERERGIGLGNYPHATENPGTSSKGLQITTTAAQIAKWEVSAIHTS 900
 Db 841 LDSRSSEKORSLEERERGIGLGNYPHATENPGTSSKGLQITTTAAQIAKWEVSAIHTS 900
 QY 901 QEDRSSGTTTELHCVTDERNALRSSAAHSTNFTKSENSENRTCSMPYAKLEYKRSS 960
 Db 901 QEDRSSGTTTELHCVTDERNALRSSAAHSTNFTKSENSENRTCSMPYAKLEYKRSS 960
 QY 961 NDSINVSSSDGYGKQGMKPSIESYSEDESDESKFCYGOYPADLAHKIHSANHMDNDGE 1020
 Db 961 NDSINVSSSDGYGKQGMKPSIESYSEDESDESKFCYGOYPADLAHKIHSANHMDNDGE 1020
 QY 1021 LDTPINYSKYDEQLNSGRQSPQNERWARPKHIIIEIKOSEQRQSNQSTTYVYTE 1080
 Db 1021 LDTPINYSKYDEQLNSGRQSPQNERWARPKHIIIEIKOSEQRQSNQSTTYVYTE 1080
 QY 1081 STDDKHLKQPHGQCEVSPYRSRGANGSETNRVGSNHNINONVSQSCICEQDDYEDDKP 1140
 Db 1081 STDDKHLKQPHGQCEVSPYRSRGANGSETNRVGSNHNINONVSQSCICEQDDYEDDKP 1140
 QY 1141 TNYSERYSSEEEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFPSKS 1200
 Db 1141 TNYSERYSSEEEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFPSKS 1200

1201 SSGSSKTEHSSSENSTSPSSNAKRONQLHPSQAQSRSGQPKAATCKVKSINQETIQ 1260
 Db 1201 SSGSSKTEHSSSENSTSPSSNAKRONQLHPSQAQSRSGQPKAATCKVKSINQETIQ 1260
 Qy 1261 TYCVEDTPICFSRCSSLSLSAEDIEGICNQTQOADSANTLQIABIKIGTRSAEDPV 1320
 Db 1261 TYCVEDTPICFSRCSSLSLSAEDIEGICNQTQOADSANTLQIABIKIGTRSAEDPV 1320
 Qy 1321 SEYPAVSOCHPRKTSRLQGSLSSEBARKHKAFFSSGAKSPSKGAQTPKSPPEHVQET 1380
 Db 1321 SEYPAVSOCHPRKTSRLQGSLSSEBARKHKAFFSSGAKSPSKGAQTPKSPPEHVQET 1380
 Qy 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGQTMPSPRSKTPP 1440
 Db 1381 PLMFSRCTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLDPSPGQTMPSPRSKTPP 1440
 Qy 1441 PPQTAQTKREVKNKAPTAKEBESGPKQAANVAQVQVLPDADTLHFAESTPDGF 1500
 Db 1441 PPQTAQTKREVKNKAPTAKEBESGPKQAANVAQVQVLPDADTLHFAESTPDGF 1500
 Qy 1501 SCSSSLSALSILDEPFIQKVELRIMPVOENDNGNETESEQPKESNENQEKAEKTIIDSE 1560
 Db 1501 SCSSSLSALSILDEPFIQKVELRIMPVOENDNGNETESEQPKESNENQEKAEKTIIDSE 1560
 Qy 1561 KDLLDDSDDDDIIELEECIIISAMPTKSRKAKPAQATASKLPPVARKPSQLPVYKLLPS 1620
 Db 1561 KDLLDDSDDDDIIELEECIIISAMPTKSRKAKPAQATASKLPPVARKPSQLPVYKLLPS 1620
 Qy 1621 QNRLOPKHVSFTPGDDMPVYCVGTPINFSTATSLSDLTIESPNELAAEGVGRGGAQ 1680
 Db 1621 QNRLOPKHVSFTPGDDMPVYCVGTPINFSTATSLSDLTIESPNELAAEGVGRGGAQ 1680
 Qy 1681 SGFEKEDTITPTEGRSTDEAOGKTSVTPPELDDNKAEEGDILAEICINSAMPKGSHPK 1740
 Db 1681 SGFEKEDTITPTEGRSTDEAOGKTSVTPPELDDNKAEEGDILAEICINSAMPKGSHPK 1740
 Qy 1741 FRVKKINDVQOASASSAENKQLOGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800
 Db 1741 FRVKKINDVQOASASSAENKQLOGKKKPTSPVKPIPONTEYTRVRKNADSKNNLN 1800
 Qy 1801 AERFSDNKKQKQNLKNNKDFNDKLPNNEDVRGSGFAFDSPPHYTPIEGTPYCFSRND 1860
 Db 1801 AERFSDNKKQKQNLKNNKDFNDKLPNNEDVRGSGFAFDSPPHYTPIEGTPYCFSRND 1860
 Qy 1861 SLSLDPDDDDVLSREKAEKLRKAKENKSEAKVTSHTELTSNQOANKTQAIKOPINR 1920
 Db 1861 SLSLDPDDDDVLSREKAEKLRKAKENKSEAKVTSHTELTSNQOANKTQAIKOPINR 1920
 Qy 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFALIENTPVCFSHNSLSLSLSDIDQENN 1980
 Db 1921 GQPKPILQKQSTFPQSSKDIIPDRGAATDEKLQNFALIENTPVCFSHNSLSLSLSDIDQENN 1980
 Qy 1981 NKNEPIKETEPDPSQSEPKQASGAPKSFHVEDTPVCFNSLSLSLSLSDIDSEDDLLQ 2040
 Db 1981 NKNEPIKETEPDPSQSEPKQASGAPKSFHVEDTPVCFNSLSLSLSLSDIDSEDDLLQ 2040
 Qy 2041 ECISAMPKPKKPSRLKGDNEKHSRPMNGGILGEDITLDDKIDQRPDSHGLSPDSENF 2100
 Db 2041 ECISAMPKPKKPSRLKGDNEKHSRPMNGGILGEDITLDDKIDQRPDSHGLSPDSENF 2100
 Qy 2101 WKAIQEGANSIVSLHQAACCLRSQAASDSDSILSLKSGISLGSFPFHLTPDQSEKPT 2160
 Db 2101 WKAIQEGANSIVSLHQAACCLRSQAASDSDSILSLKSGISLGSFPFHLTPDQSEKPT 2160
 Qy 2161 SNKGPRILKXGKSTLTETKIESGKIGKGVKYSLLITGKVRNSSEISGQMKQPLQAN 2220
 Db 2161 SNKGPRILKXGKSTLTETKIESGKIGKGVKYSLLITGKVRNSSEISGQMKQPLQAN 2220
 Qy 2221 MPGISRGRTMIHIPGVNNSSTSPVSKGPPPLKTPASKSPSEGTATTSPRGAKPSVKS 2280
 Db 2221 MPGISRGRTMIHIPGVNNSSTSPVSKGPPPLKTPASKSPSEGTATTSPRGAKPSVKS 2280
 Qy 2281 ELSPVARTSQIGSSKAPSRSGSRDTPSPRAQQLSRPIQSPGRNSISPGENGISPPN 2340

2281 ELSPVARTSQIGSSKAPSRSGSRDTPSPRAQQLSRPIQSPGRNSISPGENGISPPN 2340
 Qy 2341 KLSQIPRTSSPSTASTKSSSGSKMSYTPSGQMSQNLTKQTGLSKNASSIPRSESASKG 2400
 Db 2341 KLSQIPRTSSPSTASTKSSSGSKMSYTPSGQMSQNLTKQTGLSKNASSIPRSESASKG 2400
 Qy 2401 LNQWNGNGANKVELSRMSTKSSGSESRSRSEPVLRQSTFIKEAPSTPLRRKLEESA 2460
 Db 2401 LNQWNGNGANKVELSRMSTKSSGSESRSRSEPVLRQSTFIKEAPSTPLRRKLEESA 2460
 Qy 2461 SFESLSRSRSPASTRQAQTPVLSPLPMKSLSTHSSVQAGWRKLPPLSLPTIENDG 2520
 Db 2461 SFESLSRSRSPASTRQAQTPVLSPLPMKSLSTHSSVQAGWRKLPPLSLPTIENDG 2520
 Qy 2521 RPAKRHDIAHSHSPSRPLPINRSGTWKRSHSSSLPRVSTWRTGSSSSILSASSES 2580
 Db 2521 RPAKRHDIAHSHSPSRPLPINRSGTWKRSHSSSLPRVSTWRTGSSSSILSASSES 2580
 Qy 2581 SEKAKSDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 Db 2581 SEKAKSDEKHVNSISGTSKQSKENQVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 Qy 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPSGRSPTGNTPPVIDSVSEKANPNIKSDKN 2700
 Db 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPSGRSPTGNTPPVIDSVSEKANPNIKSDKN 2700
 Qy 2701 QAKQVNGSGVPMRTVGLNRLNSFIQVADAPDQKGTIKFGQNNPVPVSETNESSIVERT 2760
 Db 2701 QAKQVNGSGVPMRTVGLNRLNSFIQVADAPDQKGTIKFGQNNPVPVSETNESSIVERT 2760
 Qy 2761 PFSSSSSKHSPSGTVAARVTPENYNPSRKSADSTARSPOIPTVANNTKRDSKT 2820
 Db 2761 PFSSSSSKHSPSGTVAARVTPENYNPSRKSADSTARSPOIPTVANNTKRDSKT 2820
 Qy 2821 DSTESEGTQSPKSHSGSYLVTSV 2843
 Db 2821 DSTESEGTQSPKSHSGSYLVTSV 2843
 RESULT 5
 ID AAW76821 standard; protein; 2973 AA.
 AC AAW76821;
 DT 25-JAN-1999 (first entry)
 DE Human APC protein.
 KW Tcf; beta-catenin; human; drug; familial adenomatous polyposis; FAP;
 KW cancer; adenomatous polyposis coli; APC; neoplastic.
 OS Homo sapiens.
 PN WO9841631-A2.
 PD 24-SEP-1998.
 PF 20-MAR-1998; 98WO-US005506.
 PR 20-MAR-1997; 97US-00821355.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UYUT-) RIJKSUNIV UTRECHT.
 PI Barker N, Clevers H, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
 PI Vogelstein B;
 WIPI; 1998-531569/45.
 PT Intron-free DNA encoding Tcf-4 protein - useful for, e.g. identifying
 drugs for treating FAP patients, or patients with increased risk of

developing cancer.

Disclosure; Page 32-37; 58pp; English.

This sequence represents a human APC protein which is used in a method of identifying candidate drugs for use in familial adenomatous polyposis (FAP) patients, or patients with increased risk of developing cancer. The protein can also be used to determine the presence or absence in a cell of wild type adenomatous polyposis coli (APC) gene or a downstream protein in the APC transcription regulatory pathway. This method involves introducing a Tcf-responsive reporter gene into the cell, and measuring transcription of the reporter gene where a cell which supports active transcription of the reporter gene does not have wild type APC or a downstream protein of the APC transcription regulatory pathway. The protein can also be used in a method of diagnosing cancer in a sample suspected of being neoplastic and for treating a patient with colorectal cancer or other cancer associated with FAP, comprising administering to the patient a polypeptide comprising a portion of the APC sequence, comprising the beta-catenin binding site

Sequence 2973 AA;

Query Match 100.0%; Score 14566; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAAYDQLLKQVEALKXENSLRQELNDSNHLTKLETEASNKXEVKQLQGSIEDAM 60
DB 1 MAAAYDQLLKQVEALKXENSLRQELNDSNHLTKLETEASNKXEVKQLQGSIEDAM 60
QY 61 ASSGOIDLLERLKLNLSSNPPGVKLRKMSLRYSRGREGSVSRSGECSPVPMGSPPR 120
DB 61 ASSGOIDLLERLKLNLSSNPPGVKLRKMSLRYSRGREGSVSRSGECSPVPMGSPPR 120
QY 121 RGVNGSRSTGYLBELEKERSLLADLDKEEKXDYQAOQLTKRIDLPLTEPNSL 180
DB 121 RGVNGSRSTGYLBELEKERSLLADLDKEEKXDYQAOQLTKRIDLPLTEPNSL 180
QY 181 QDTMTROLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQIEKIDILRQLQSQAT 240
DB 181 QDTMTROLEYEARQIRVAMEEQLGTCQDMKRAQRRIARIQIEKIDILRQLQSQAT 240
QY 241 EAERSSQNKHTGSHDAERQEGQVGEINMATSGOGSTTRMDHETASVLSSTHSA 300
DB 241 EAERSSQNKHTGSHDAERQEGQVGEINMATSGOGSTTRMDHETASVLSSTHSA 300
QY 301 PRRLTSHLGTVMYVYLLSMLGTHDKDDMSRTLLIAMSSQDSCISMRQSGCLPLLIQLL 360
DB 301 PRRLTSHLGTVMYVYLLSMLGTHDKDDMSRTLLIAMSSQDSCISMRQSGCLPLLIQLL 360
QY 361 HGNDKSVLLGNRSGSKEARARASALHNIHISQDDKGRREIRVLHLEQIRAYCETC 420
DB 361 HGNDKSVLLGNRSGSKEARARASALHNIHISQDDKGRREIRVLHLEQIRAYCETC 420
QY 421 WEQEAHEFGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRRHANNELGLOAIELLO 480
DB 421 WEQEAHEFGMDQKNPMPAPVEHQICPAVCVLMKLSFDEHRRHANNELGLOAIELLO 480
QY 481 VDCMYGLTNDHYISITLRRYAGMALTNLTTFQDVANKATLCSMKCMALVAQLKSEEDL 540
DB 481 VDCMYGLTNDHYISITLRRYAGMALTNLTTFQDVANKATLCSMKCMALVAQLKSEEDL 540
QY 541 QQVIASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKVLSALMNLISACT 600
DB 541 QQVIASVLRLNLSRADVNSKTLREVGSVKALMECALEVKKESTLKVLSALMNLISACT 600
QY 601 ENKADICAVDGALAFVGLTYRSTQNTLAIIESGGGILRVSSLIATNEDHRIQLRENN 660
DB 601 ENKADICAVDGALAFVGLTYRSTQNTLAIIESGGGILRVSSLIATNEDHRIQLRENN 660
QY 661 CLOTLLOHLKSHSITIVSNAGCTLWNLISARNPKDQBALMDGAVSMKLNLIHSHKGMIA 720
DB 661 CLOTLLOHLKSHSITIVSNAGCTLWNLISARNPKDQBALMDGAVSMKLNLIHSHKGMIA 720

QY 721 GSAALRNLMANRPKAYKNDANIMSPGSSLPSELHVRKQKALAEALDAQHLSETFDNIDNLS 780
DB 721 GSAALRNLMANRPKAYKNDANIMSPGSSLPSELHVRKQKALAEALDAQHLSETFDNIDNLS 780
QY 781 PKASHRSKQRHKQSLYGDYVFTNRHDDNDSNPNTGMTVLSPLYLNTTVLPSSSSSRGS 840
DB 781 PKASHRSKQRHKQSLYGDYVFTNRHDDNDSNPNTGMTVLSPLYLNTTVLPSSSSSRGS 840
QY 841 LDSRSSEKDSRLERBERGIGLGNYPHATENPQTSKRGIGLQISTTAAQIAKMWEEVSAIHTS 900
DB 841 LDSRSSEKDSRLERBERGIGLGNYPHATENPQTSKRGIGLQISTTAAQIAKMWEEVSAIHTS 900
QY 901 QEDSSSGTTELHCVTDERNALRRSSAAHNTNTYNTKSENSNRTCSMPYAKLEYKRSS 960
DB 901 QEDSSSGTTELHCVTDERNALRRSSAAHNTNTYNTKSENSNRTCSMPYAKLEYKRSS 960
QY 961 NDSLSNVSSSDGYKRGOMKPSIESYEDDESFCYSYQYQPADLAHAKIHANEMDDNDGE 1020
DB 961 NDSLSNVSSSDGYKRGOMKPSIESYEDDESFCYSYQYQPADLAHAKIHANEMDDNDGE 1020
QY 1021 LDTPIYSLKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTTPYVTE 1080
DB 1021 LDTPIYSLKYSDQLNSGRQSPQNERWARPKHIIIEDEIKQSEQRQSRNQSTTTPYVTE 1080
QY 1081 STDDKHLKFPQHFQGCVCVPSYSRGANGSETNRVGSNHGINQVNSQSLCOEDDYEDDKP 1140
DB 1081 STDDKHLKFPQHFQGCVCVPSYSRGANGSETNRVGSNHGINQVNSQSLCOEDDYEDDKP 1140
QY 1141 TNSYRYSEBEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFSPSKS 1200
DB 1141 TNSYRYSEBEQHEEERPTNYSIKYNEEKRVDPQIDYSLKYATDIPSSQKQSFSPSKS 1200
QY 1201 SSGOSSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRGQPKAATCKVSSINQETIQ 1260
DB 1201 SSGOSSKTEHMSSESSENTSTPSSNAKRONQLHPSSAQSRGQPKAATCKVSSINQETIQ 1260
QY 1261 TYCVEDTPICFSCSSLSLSAEDTGCNQTQEADSNANTLOIAIEKIKIGTRSAEDPV 1320
DB 1261 TYCVEDTPICFSCSSLSLSAEDTGCNQTQEADSNANTLOIAIEKIKIGTRSAEDPV 1320
QY 1321 SEVPAYSQHPRTKSKLQSSLSSESARHKAFFSSGAKSPSKSGAQTPKSPHYVQET 1380
DB 1321 SEVPAYSQHPRTKSKLQSSLSSESARHKAFFSSGAKSPSKSGAQTPKSPHYVQET 1380
QY 1381 PLMFSCSTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPPSRSKTTP 1440
DB 1381 PLMFSCSTSVSSLDSPESRSIASSVQSEPCSGMVSGIISPSDLPDPSGQTMPPSRSKTTP 1440
QY 1441 PPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLHLHFATESPDGF 1500
DB 1441 PPQTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLHLHFATESPDGF 1500
QY 1501 SCSSSLSALSLDEPFTOKDVELRIMPVQENNGNETSEQPKESNENQEKAEKTIIDSE 1560
DB 1501 SCSSSLSALSLDEPFTOKDVELRIMPVQENNGNETSEQPKESNENQEKAEKTIIDSE 1560
QY 1561 KOLLDDSDDDDIIELEECIIISAMPTKSSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
DB 1561 KOLLDDSDDDDIIELEECIIISAMPTKSSRKAKKPAQTASKLPPPVARKPSQLPVYKLLPS 1620
QY 1621 QNRLOPKHVSTPGDDMPRVYCVBGTPIINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
DB 1621 QNRLOPKHVSTPGDDMPRVYCVBGTPIINFSTATSLDLTIESPPNELAAGEVGRGAQ 1680
QY 1681 SGBEFEXKRTIPIPEGSTDEAOGKTSVVITPELDDNKAEEGDILAEICINSAMPKGSHPK 1740
DB 1681 SGBEFEXKRTIPIPEGSTDEAOGKTSVVITPELDDNKAEEGDILAEICINSAMPKGSHPK 1740
QY 1741 FRVKKIMDVOQASASSAPNKNQLDGKKKCTSPVKPIQNTYRTRVRKNADSKNNLN 1800
DB 1741 FRVKKIMDVOQASASSAPNKNQLDGKKKCTSPVKPIQNTYRTRVRKNADSKNNLN 1800

QY	1801	ABRVFSDNKKDKKQNLKNNKDFNDKLPNNEDVRGSAFADSPHYHTPIEGTPYCFSRND	1860
DB	1801	ABRVFSDNKKDKKQNLKNNKDFNDKLPNNEDVRGSAFADSPHYHTPIEGTPYCFSRND	1860
QY	1861	SLSLDFDDDDVDLREKAELEKAKENKESAKVTSHLTSTNOQSANKTQAIKQPINR	1920
DB	1861	SLSLDFDDDDVDLREKAELEKAKENKESAKVTSHLTSTNOQSANKTQAIKQPINR	1920
QY	1921	GQPKILOQSTPPOSSKDIPIRGAAATDEKLNFAIENTPVCFSNSSLSSLSIDIQENN	1980
DB	1921	GQPKILOQSTPPOSSKDIPIRGAAATDEKLNFAIENTPVCFSNSSLSSLSIDIQENN	1980
QY	1981	NKNEPIKTEPPDPSGQSPKQASGYAPKSFHVEDTPVCFERNSSLSSLSIDSDLLQ	2040
DB	1981	NKNEPIKTEPPDPSGQSPKQASGYAPKSFHVEDTPVCFERNSSLSSLSIDSDLLQ	2040
QY	2041	ECISSAMPKPKPSRLKGDNEKHSPRNMGILGEDLTDLKDIQRPDSEHGLSPDSENF	2100
DB	2041	ECISSAMPKPKPSRLKGDNEKHSPRNMGILGEDLTDLKDIQRPDSEHGLSPDSENF	2100
QY	2101	WKAIQEGANSIVSLHQAACAAACLSQAQSDSDSILSLXSGISLGSPPHLLTPDQEKPT	2160
DB	2101	WKAIQEGANSIVSLHQAACAAACLSQAQSDSDSILSLXSGISLGSPPHLLTPDQEKPT	2160
QY	2161	SNKGPRILKPEKSTLETTKIESGKIGKGVKYSLLITGKVRNSEISGQWKQLOAN	2220
DB	2161	SNKGPRILKPEKSTLETTKIESGKIGKGVKYSLLITGKVRNSEISGQWKQLOAN	2220
QY	2221	MPSISRGRTMIHIPGVNRNSSLSPSVSKGPPKLTTPASKSPSEGQATTSPRGAKEPVKS	2280
DB	2221	MPSISRGRTMIHIPGVNRNSSLSPSVSKGPPKLTTPASKSPSEGQATTSPRGAKEPVKS	2280
QY	2281	ELSPVARTQSIGGSSKAPRSGSRDTPSRPAQPLSRPIQSPGRNSISPGNGISPPN	2340
DB	2281	ELSPVARTQSIGGSSKAPRSGSRDTPSRPAQPLSRPIQSPGRNSISPGNGISPPN	2340
QY	2341	KLSQLPRTSPSTASTSSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG	2400
DB	2341	KLSQLPRTSPSTASTSSGSGKMSYTPSGRQMSQNLTKQTGLSKNASSIPRSESASKG	2400
QY	2401	LNQMNNGANKVELSRMSTKSSGESDRSRPVLVROSTTIKEAPSTLRKLEESA	2460
DB	2401	LNQMNNGANKVELSRMSTKSSGESDRSRPVLVROSTTIKEAPSTLRKLEESA	2460
QY	2461	SFESLSPSRPASPTRSQAOQTVLPSLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG	2520
DB	2461	SFESLSPSRPASPTRSQAOQTVLPSLPDMSLSTHSSVQAGWRKLPPLNLSPTIEYNDG	2520
QY	2521	RPAKRHDIAHSHSPSRPLNIRSGTWKREHSHSSLPVSVTWRTTGGSSSILSASSES	2580
DB	2521	RPAKRHDIAHSHSPSRPLNIRSGTWKREHSHSSLPVSVTWRTTGGSSSILSASSES	2580
QY	2581	SEKAKSEDEKVNISIGTKQSKENQVAKGTWRIKENEPSTNISTQTVSSGATNGAES	2640
DB	2581	SEKAKSEDEKVNISIGTKQSKENQVAKGTWRIKENEPSTNISTQTVSSGATNGAES	2640
QY	2641	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPTGNTFPVIDSVSEKANPNIKSKON	2700
DB	2641	KTLIYQMAPAVSKTEDVWVRIEDCPINNPRSGRSPTGNTFPVIDSVSEKANPNIKSKON	2700
QY	2701	QAKQNVGNGVPMRTVGLNENLNFIOVDAPDQGTIKQNNPVPVSTNESSIIVERT	2760
DB	2701	QAKQNVGNGVPMRTVGLNENLNFIOVDAPDQGTIKQNNPVPVSTNESSIIVERT	2760
QY	2761	PFSSSSSSKHSSPSGTVAARVTPFNYPNPSPKSSADSTSAEPSQIPTPVNNTKKRDSKT	2820
DB	2761	PFSSSSSSKHSSPSGTVAARVTPFNYPNPSPKSSADSTSAEPSQIPTPVNNTKKRDSKT	2820
QY	2821	DSTESSGTSQPKRHSGLYLTSV	2843
DB	2821	DSTESSGTSQPKRHSGLYLTSV	2843

RESULT 6	
AAV72782	standard; protein; 2973 AA.
ID	AAV72782
XX	AAV72782;
AC	AAV72782;
XX	31-MAY-2001 (first entry)
DT	31-MAY-2001 (first entry)
XX	Transcriptional activation protein #1 related to the invention.
DE	Human; Tcf-4 protein; transcriptional activation factor; beta-catenin;
XX	apoptosis; colorectal cancer; Familial Adenomatous Polyposis; FAP; APC;
KW	adenomatous polyposis coli; recombinant adenovirus; Ad-Mini-Me; therapy;
KW	GFP; green fluorescent protein; GFP/cAPC fusion protein; cytostatic.
XX	Homo sapiens.
OS	WO200116167-A2.
XX	08-MAR-2001.
PN	29-AUG-2000; 2000WO-US023635.
XX	01-SEP-1999; 99US-00388354.
PD	(UYJO) UNIV JOHNS HOPKINS.
XX	Barker N, Clevers JC, Kinzler KW, Korinek V, Morin PJ, Sparks AB;
PI	Vogelstein B, He T;
PI	WPI; 2001-226675/23.
XX	New fusion protein comprising an enzyme covalently linked to a portion of
PT	the adenomatous polyposis coli comprising its beta-catenin binding domain
PT	useful for inducing apoptosis or treating colorectal cancer.
XX	Disclosure; Page 74-81; 83pp; English.
FS	The present sequence is a transcriptional activation protein related to
XX	the invention. The invention relates to human Tcf-4 proteins and their
CC	corresponding cDNA molecules which encodes transcriptional activation
CC	factors of human Tcf/Lef family. Human Tcf-4 binds to beta-catenin and
CC	activates transcription in colorectal epithelial cells. Moreover it has
CC	been found that adenomatous polyposis coli (APC) regulates this
CC	transcriptional activation, by binding to beta-catenin. The invention
CC	also provides a recombinant adenovirus, Ad-Mini-Me i.e., APC Minus its
CC	amino- and carboxyl-terminal Ends which expresses a fusion protein,
CC	green fluorescent protein (GFP)/cAPC containing GFP fused to the central
CC	third of APC which contains its beta-catenin binding domain useful for
CC	inducing apoptosis or treating colorectal cancer. These fusion proteins
CC	are useful for treating cancer, e.g. colorectal cancer, and other cancers
CC	associated with Familial Adenomatous Polyposis (FAP) or patients with
CC	increased risk of developing cancer. Human Tcf-4 cDNA provides an
CC	excellent system for screening agents for their ability to promote
CC	delivery, integration, hybridization, expression, replication or
CC	integration in cells or in an animal. It also provides methods for
CC	diagnosis cancer in a sample suspected of being neoplastic
XX	
SQ	Sequence 2973 AA;
	Query Match 100.0%; Score 14566; DB 4; Length 2973;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAAASVDQLLKQVEALKMNSNLRLQELDNHNLTKLEASNMKEVLKQLQGSIEDAM	60
DB	1	MAAASVDQLLKQVEALKMNSNLRLQELDNHNLTKLEASNMKEVLKQLQGSIEDAM	60
QY	61	ASSGQIDLLERLKLNLDSNFPVGLRSMKLSRSGRSRSGSCSPVPMGSPFR	120
DB	61	ASSGQIDLLERLKLNLDSNFPVGLRSMKLSRSGRSRSGSCSPVPMGSPFR	120
QY	121	RGFVNGRSTGYLELEKERSLLADLDKEKEDWYQAQLNLTKRIDSPLTENFSL	180

121 RGFVNGSRSTGYLEBLEKERSULLADLDKEBEKDWYIAQONLTKRIDSPLTFENFSL 180
181 QIDMTROQLEYEARQIRVAMEEQQLGTCODMEKRAQRRIARIQQIEBKDILIRIQLLOQSOAT 240
181 QIDMTROQLEYEARQIRVAMEEQQLGTCODMEKRAQRRIARIQQIEBKDILIRIQLLOQSOAT 240
241 EAEERSONKHETGSHDAERONGEGVGEBINMATSGNGQSTTRMDHETASVLSSSSTHSA 300
241 EAEERSONKHETGSHDAERONGEGVGEBINMATSGNGQSTTRMDHETASVLSSSSTHSA 300
301 PRRLTSHLGTKVEMYVSYLLSMGLGTHDKDDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL 360
301 PRRLTSHLGTKVEMYVSYLLSMGLGTHDKDDMSRTLLAMSSQSDSCISMRQSGCLPLLIQLL 360
361 HGNDKDSVLLGNRSRSGKEARASAAALHNI IHQOPDDKRGREIRIVLHLLBOIRAYCETC 420
361 HGNDKDSVLLGNRSRSGKEARASAAALHNI IHQOPDDKRGREIRIVLHLLBOIRAYCETC 420
421 MEWQEAHEPGMDODKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGIQAIAELLO 480
421 MEWQEAHEPGMDODKNPMPAPVEHQICPAVCVLMKLSFDEEHRHAMNELGGIQAIAELLO 480
481 VDCEMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWALVAQLKSESEDL 540
481 VDCEMYGLTNDHYSITLRYAGMALTNLTFGDVANKATLCSMKGCWALVAQLKSESEDL 540
541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALVKKESTLKSVLNALWNLSAHCT 600
541 QOVIASVLRNLSWRADVNSKTLREVGSVKALMECALVKKESTLKSVLNALWNLSAHCT 600
601 ENKADICAVDGAFLVGLTLYRSQTNLTALIESGGILRVNVSLLIATNEDHRQILRENN 660
601 ENKADICAVDGAFLVGLTLYRSQTNLTALIESGGILRVNVSLLIATNEDHRQILRENN 660
661 CLQTLLOHLKSHSLTIVSNACOTLWNLARPQDOEALWDMGAVSMLKNLHSHKHMIA 720
661 CLQTLLOHLKSHSLTIVSNACOTLWNLARPQDOEALWDMGAVSMLKNLHSHKHMIA 720
721 GSAALRNLMANRPKAKYKADANIMSPGSSLPSLHVKKOKALAEALDQOHLSTTFNIDNLS 780
721 GSAALRNLMANRPKAKYKADANIMSPGSSLPSLHVKKOKALAEALDQOHLSTTFNIDNLS 780
781 PKASHRSQRKHQSLYGYVFDTNHRDNRDNFNTGNTVLSPLYNTTLPVSSSSSRGS 840
781 PKASHRSQRKHQSLYGYVFDTNHRDNRDNFNTGNTVLSPLYNTTLPVSSSSSRGS 840
841 LDSSRSEKDRSLERERGLGLGNYPHATENPGTSSKRGLOISTTAAQIAKWEVSAIHTS 900
841 LDSSRSEKDRSLERERGLGLGNYPHATENPGTSSKRGLOISTTAAQIAKWEVSAIHTS 900
901 QEDRSSGTTTELHCVTDERNALRSSAAHTSNVTNFTKSENSNRTCSMPYAKLEYKRSS 960
901 QEDRSSGTTTELHCVTDERNALRSSAAHTSNVTNFTKSENSNRTCSMPYAKLEYKRSS 960
961 NDSLNSVSSDGYKRGOMKPSIESYSEDDSKFCSYQOYPADLAHKHSHANWMDNDGE 1020
961 NDSLNSVSSDGYKRGOMKPSIESYSEDDSKFCSYQOYPADLAHKHSHANWMDNDGE 1020
1021 LDTPIYSLKYSDEOLNSGRQSPONERWARPKHIIIEDEIKQSEQRORSNOSTTYPVYTE 1080
1021 LDTPIYSLKYSDEOLNSGRQSPONERWARPKHIIIEDEIKQSEQRORSNOSTTYPVYTE 1080
1081 STDDKHLXFPQHFQGOECVSPYRSGANGSTNRVGSNHGINQVNSQSLCOEDDYEDDKP 1140
1081 STDDKHLXFPQHFQGOECVSPYRSGANGSTNRVGSNHGINQVNSQSLCOEDDYEDDKP 1140
1141 TNYSERYSEREEHREERTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1200
1141 TNYSERYSEREEHREERTNYSIKYNEEKHVDQPIDYSLKYATDIPSSOKQSFSEKS 1200
1201 SSGQSKTEHMSSESTSTPSSNAKRONQLHPSSAQRSGQPQKAATCKVSSINQETIQ 1260

1201 SSGQSKTEHMSSESTSTPSSNAKRONQLHPSSAQRSGQPQKAATCKVSSINQETIQ 1260
1261 TYCVEDTPICFSRCSSLSLSAASDEIGCNOTTOEADSAANTLOIAEIKELKIGTRSADPV 1320
1261 TYCVEDTPICFSRCSSLSLSAASDEIGCNOTTOEADSAANTLOIAEIKELKIGTRSADPV 1320
1321 SEVPAVQCHPRTKSRLQGSLSLSAARHKAARFSSGAQSPKSPGHHYVOET 1380
1321 SEVPAVQCHPRTKSRLQGSLSLSAARHKAARFSSGAQSPKSPGHHYVOET 1380
1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMWVGIIISPSDLPDPSPGOTMPPSRSTKTP 1440
1381 PLMFESRCTSVSSLDSPESRSIASSVQSEPCSGMWVGIIISPSDLPDPSPGOTMPPSRSTKTP 1440
1441 PPPOTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLLHFAESTPDGF 1500
1441 PPPOTAQTKREVPKNKAPTAEKRESGPKQAAVNAAVQVQVLPDADTLLHFAESTPDGF 1500
1501 SCSSLSALSILDEPFFIQKQVLELIMPPVQENDNGNETESQPKESNEQKEAEKTIIDSE 1560
1501 SCSSLSALSILDEPFFIQKQVLELIMPPVQENDNGNETESQPKESNEQKEAEKTIIDSE 1560
1561 KDLLDDDDDDIIELEECIIISAMPTKSSRKAKKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1561 KDLLDDDDDDIIELEECIIISAMPTKSSRKAKKPAQATASKLPPPVARKPSQLPVYKLLPS 1620
1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPNELAAEGVRGGAQ 1680
1621 QNRLOPKHVSFTPGDDMPRVYCEGTPINFSTATSLSDLTIESPNELAAEGVRGGAQ 1680
1681 SGFEFEKEDTTPTEGRSTDEAOGGKTSSVTIPELDDNKAEBEGDILAEICINSAMPKGSHP 1740
1681 SGFEFEKEDTTPTEGRSTDEAOGGKTSSVTIPELDDNKAEBEGDILAEICINSAMPKGSHP 1740
1741 FRVKKIMQVQOASASSAPNKNQLOGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1800
1741 FRVKKIMQVQOASASSAPNKNQLOGKKKPTSPVKPIPQNTTEYTRVRKNADSKNNLN 1800
1801 AERFESDNKDKKONLKNNSKDFENDLKNNEEDVRGSFAFDSPHHYTPIEGTTCYCSRND 1860
1801 AERFESDNKDKKONLKNNSKDFENDLKNNEEDVRGSFAFDSPHHYTPIEGTTCYCSRND 1860
1861 SLSSLPDDDDVLSREKAEELRKAKENKESAEKVTSTHTELTSNQQSANKTQAIKAPINR 1920
1861 SLSSLPDDDDVLSREKAEELRKAKENKESAEKVTSTHTELTSNQQSANKTQAIKAPINR 1920
1921 GQPKPILQKOSTTPOSSKOIPDRGAATDEKLQNFATENTPVCFSHNSLSLSDIDQENN 1980
1921 GQPKPILQKOSTTPOSSKOIPDRGAATDEKLQNFATENTPVCFSHNSLSLSDIDQENN 1980
1981 NKENEPIKETEPDPSQGBPSKPOASGYAPKSFHVEDTPVCFSRNSLSLSDIDEDLLQ 2040
1981 NKENEPIKETEPDPSQGBPSKPOASGYAPKSFHVEDTPVCFSRNSLSLSDIDEDLLQ 2040
2041 ECISAMPKPKKPSRLKGDNEKHSPPNMGIIIGEDJTLDLKOIQRDPDSEHGLSPDSENF 2100
2041 ECISAMPKPKKPSRLKGDNEKHSPPNMGIIIGEDJTLDLKOIQRDPDSEHGLSPDSENF 2100
2101 WKAIQSGANSIVSSLHQAACCLSRQASDSDSILSLKSGISLGSPEHLTPDOEEKPFT 2160
2101 WKAIQSGANSIVSSLHQAACCLSRQASDSDSILSLKSGISLGSPEHLTPDOEEKPFT 2160
2161 SNKGPRILKPEKSTLETTKKIESEKGIKGGKKVYKSLITGKVRNSNSEISGQWKQPLQAN 2220
2161 SNKGPRILKPEKSTLETTKKIESEKGIKGGKKVYKSLITGKVRNSNSEISGQWKQPLQAN 2220
2221 MPSISRGRTMIHLPVGRNSSSTSPVSKGPKPLTKPASKSPSEGQTATTPRGAKPSVKS 2280
2221 MPSISRGRTMIHLPVGRNSSSTSPVSKGPKPLTKPASKSPSEGQTATTPRGAKPSVKS 2280
2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQQLSRPIQSPGRNSISPGRNGISPPN 2340
2281 ELSPVARQTSQIGGSSKAPSRGSRDSTPSRPAQQLSRPIQSPGRNSISPGRNGISPPN 2340

2341 KLSQLPRTSPSTASTKSSGSGKMSYTSFGQMSQNLTKOTGLSKNASSIPRSESASKG 2400
 2341 KLSQLPRTSPSTASTKSSGSGKMSYTSFGQMSQNLTKOTGLSKNASSIPRSESASKG 2400
 2401 LNMNNGANKVVELSRMSSTKSSGSDRPERPVLVQSTFIKEAPSTLRKLESA 2460
 2401 LNMNNGANKVVELSRMSSTKSSGSDRPERPVLVQSTFIKEAPSTLRKLESA 2460
 2461 SFESLSPSRPASPTSQACTPVLSPDMSLSTHSSVQAGGWRKLPNLSPTIYNDG 2520
 2461 SFESLSPSRPASPTSQACTPVLSPDMSLSTHSSVQAGGWRKLPNLSPTIYNDG 2520
 2521 RPAKRHDIAHSHESPSRLPINRSGTWKREHSHKSSSLPRVSTWRTTSGSSILSAGES 2580
 2521 RPAKRHDIAHSHESPSRLPINRSGTWKREHSHKSSSLPRVSTWRTTSGSSILSAGES 2580
 2581 SEKADEDEKHVNSISGTQSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 2581 SEKADEDEKHVNSISGTQSKENOVSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAES 2640
 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPRSGSPGTGTPPVIDSVEKANPNIKDKDN 2700
 2641 KTLIYQMAPAVSKTEDVWRIEDCPINNPRSGSPGTGTPPVIDSVEKANPNIKDKDN 2700
 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVDAPDOKGTEIKPGQNNPVVSETNESSIVERT 2760
 2701 QAKQNVGSGVPMRTVGLNRLNSFIQVDAPDOKGTEIKPGQNNPVVSETNESSIVERT 2760
 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPSRKSSADSTGARPQIPTPVNNNTKEDSKT 2820
 2761 PFSSSSSKHSSPSGTVAAARVTPFNPNPSRKSSADSTGARPQIPTPVNNNTKEDSKT 2820
 2821 DSTESSGTQSPKRGSGSYLTVSV 2843
 2821 DSTESSGTQSPKRGSGSYLTVSV 2843

RESULT 7

AA170304

ID AAY70304 standard; protein; 2973 AA.

AC AAY70304;

DT 06-JUN-2000 (first entry)

XX Protein used in cancer diagnosis associated with APC/beta-catenin.

XX Human Tcf-4E protein; transcription factor; beta-catenin;
 KW Adenomatous polyposis coli; APC; transcriptional activator;
 KW Tcf responsive reporter gene; APC transcription regulatory pathway;
 KW familial adenomatous polyposis; FAP; cancer; colorectal; thyroid; brain;
 KW medulloblastoma; breast; head; neck; desmoid tumour; osteoma; cytostatic.

XX Homo sapiens.

XX WO200011195-A1.

XX 02-MAR-2000.

XX 20-AUG-1999; 99WO-US018774.

XX 20-AUG-1998; 98US-00136605.

XX (UJJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX He T, Vogelstein B, Kinzler KW;

XX WPI; 2000-237657/20.

XX Determining wild-type adenomatous polyposis coli protein for diagnosing
 PT cancer comprises introducing a Tcf responsive reporter gene having
 PT upstream sequences of c-Myc into a cell.

XX Disclosure; Page 58-65; 70pp; English.
 XX The present sequence is a protein involved in cancer diagnosis associated
 CC with APC or beta-catenin mutations. Human Tcf-4E protein expressed in
 CC colorectal epithelium transactivates transcription when associated with
 CC beta-catenin. Adenomatous polyposis coli (APC) regulates this
 CC transcriptional activation, at least in part by binding to beta-catenin.
 CC Determining wild-type APC protein for diagnosing cancer comprises
 CC introducing a Tcf responsive reporter gene having upstream sequences of c
 CC -Myc into a cell and measuring transcription of the reporter gene. The
 CC candidate drug identified is useful for treating familial adenomatous
 CC polyposis patients with APC or beta-catenin mutations and patients with
 CC increased risk of developing cancers such as colorectal, thyroid, brain,
 CC medulloblastoma, desmoid tumour, osteoma, breast, head and neck
 XX Sequence 2973 AA;

Query Match 100.0%; Score 14559; DB 3; Length 2973;
 Best Local Similarity 100.0%; P-red. No. 0;
 Matches 2842; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAASYPQLLKQVEALKWENSINRQELSDNSHLTKLETEASNKKEVLKQLQSIEDEAM 60
 DB 1 MAASYPQLLKQVEALKWENSINRQELSDNSHLTKLETEASNKKEVLKQLQSIEDEAM 60
 QY 61 ASSGOIDLLERLKLNDLSNFGVGLRSKMSLRSYSGREGSVSSRGSGECPVPMGSFPR 120
 DB 61 ASSGOIDLLERLKLNDLSNFGVGLRSKMSLRSYSGREGSVSSRGSGECPVPMGSFPR 120
 QY 121 RGFVNGRESTGYLEELKERSLLADLDEEKEKQWYIAQLNLTKRISLPLENFSL 180
 DB 121 RGFVNGRESTGYLEELKERSLLADLDEEKEKQWYIAQLNLTKRISLPLENFSL 180
 QY 181 QDTMTRRQLEYEARQIRVAMEEQLGTQDMKRAQRRIARIQIEKDIILRIQLQSOAT 240
 DB 181 QDTMTRRQLEYEARQIRVAMEEQLGTQDMKRAQRRIARIQIEKDIILRIQLQSOAT 240
 QY 241 EAERSQNKHETGSHDAERQNGGVEINMATSNGCGGSTRMDHETASVLSSTHSA 300
 DB 241 EAERSQNKHETGSHDAERQNGGVEINMATSNGCGGSTRMDHETASVLSSTHSA 300
 QY 301 PRRLTSHLGTKEVWYISLLMLGTHDKDMSRTLLAMSSSDSCISMRQSGCLPLLQLL 360
 DB 301 PRRLTSHLGTKEVWYISLLMLGTHDKDMSRTLLAMSSSDSCISMRQSGCLPLLQLL 360
 QY 361 HGNDKDSVLLGNRSGSKARARASAAHLNITHSQPDDKRGRRIRVHLLEQIRAYCETC 420
 DB 361 HGNDKDSVLLGNRSGSKARARASAAHLNITHSQPDDKRGRRIRVHLLEQIRAYCETC 420
 QY 421 WEQBAHPGMDQKNPMPAVEHQICPAVCVLMKLSFDEHRRHAMELGLOIAELLQ 480
 DB 421 WEQBAHPGMDQKNPMPAVEHQICPAVCVLMKLSFDEHRRHAMELGLOIAELLQ 480
 QY 481 VDCENYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 540
 DB 481 VDCENYGLTNDHYSITLRRYAGMALTNLTFGDVANKATLCMKGCMRALVAQLKSEEDL 540
 QY 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALNLSAHT 600
 DB 541 QQVIASVLRNLSWRADVNSKTLREVGSVKALMECALEVKESTLKSVALNLSAHT 600
 QY 601 ENKADICAVDGLAFLVGTLTYSQTNTLAIIESGGILRVNSSLIATNEDHRIQIRENN 660
 DB 601 ENKADICAVDGLAFLVGTLTYSQTNTLAIIESGGILRVNSSLIATNEDHRIQIRENN 660
 QY 661 CLQTLLOHLKSHSLTIVSNACGTLINLSARPKNQOEALWDMGAVSMKLNLSHSHKMIAM 720
 DB 661 CLQTLLOHLKSHSLTIVSNACGTLINLSARPKNQOEALWDMGAVSMKLNLSHSHKMIAM 720
 QY 721 GSAARLRNLMANRPKYKDANIMSPGSSLPISHVRKQKALEAELDAQHLSETFDNIDLS 780
 DB 721 GSAARLRNLMANRPKYKDANIMSPGSSLPISHVRKQKALEAELDAQHLSETFDNIDLS 780